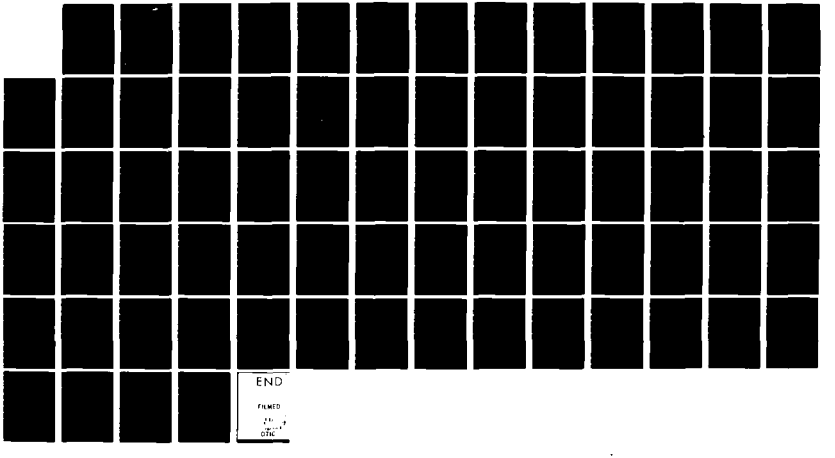
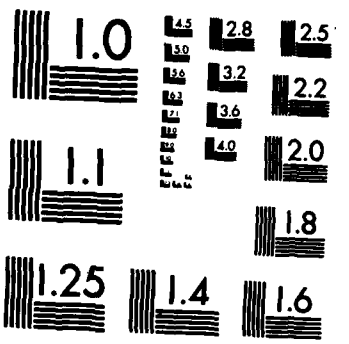
AD-A125 869 INVESTIGATE AND DEVELOP MATHEMATICAL FORMULAS IN SUPPORT OF ATMOSPHERIC STUDIES(U) BEDFORD RESEARCH ASSOCIATES MA J NOONAN ET AL. 15 JAN 83 1/1

UNCLASSIFIED AFGL-TR-83-0003 F19628-79-C-0163 F/G 4/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A125869

AFGL-TR-83-0003

INVESTIGATE AND DEVELOP MATHEMATICAL FORMULAS IN SUPPORT OF ATMOSPHERIC STUDIES

J. Noonan
D. Dechichio
K. Scharr
N. Scotti
S. Chua
P. Meehan
H. Wadzinski

Bedford Research Associates
4 DeAngelo Drive
Bedford, Massachusetts 01730

Final Report
August 1979 - September 1982

15 January 1983

DTIC FILE COPY

Approved for public release; distribution unlimited

DTIC
ELECTE
MAR 2 1 1983
A

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

83 03 21 021

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFGL-TR-83-0003	2. GOVT ACCESSION NO.: AD-A125869
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Investigate and Develop Mathematical Formulas in Support of Atmospheric Studies	5. TYPE OF REPORT & PERIOD COVERED: FINAL REPORT August 1979 - September 1982
	6. PERFORMING ORG. REPORT NUMBER: FINAL REPORT
7. AUTHOR(s): J. Noonan, N. Scotti, P. Meehan, D. Dechichio, S. Chua, H. Wadzinski, K. Scharr	8. CONTRACT OR GRANT NUMBER(s): F19628-79-C-0163
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Bedford Research Associates, 4 De Angelo Drive, Bedford, MA 01730	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62101F, 9993XXXX
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Geophysics Laboratories, Hanscom AFB, MA 01731, Monitor: Paul Tsipouras/SUNA	12. REPORT DATE: 15 January 1983
	13. NUMBER OF PAGES: 73
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Magnetometer Network, Atmospheric Turbulence, Ionospheric Ion Densities, Electron and photon cross sections

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents a summary of analyses and mathematical modelling done in support of some AFGL atmospheric research projects. In particular, work on the problems of atmospheric turbulence, magnetometer network, DMSP Ion density modelling, and Low Energy Electron and Photon cross section studies are discussed.

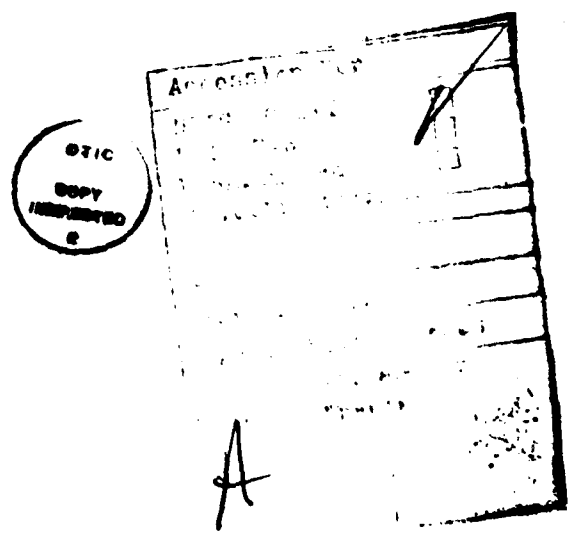
DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

	Page
Software Development to Support the 2nd AFGL Geomagnetism Workshop	5
Minuteman Data Analysis	24
Terrain Effects on Turbulence Model	35
A Technique for Estimating Uniform Grids of Ion Density from Irregularly Spaced Data	51
Low Energy Electron and Photon Cross Sections for O, N_2, and O_2 and Related Data	70



SOFTWARE DEVELOPMENT TO SUPPORT THE 2ND APGL GEOMAGNETISM WORKSHOP

1. INTRODUCTION

1.1 Overview

This report describes the technical and software procedures provided by Bedford Research Associates for the Geomagnetism Workshop on the topic "Impulse Response of the Magnetosphere" that was held at the Air Force Geophysics Laboratory (APGL) at Hanscom Air Force Base from December 2, 1981 to December 4, 1981.

1.2 Technical Background

The purpose of this workshop was to study a single "SSC" (storm sudden commencement) with information obtained from as many magnetometer networks as possible. An SSC results when a flare or other explosive event on the sun generates an interplanetary travelling shock wave (a blast). This shock propagates at speeds in the order of 100 km/sec and arrives in the vicinity of earth about 40 hours later. When the shock strikes the magnetopause, the result is a sudden compression of the magnetosphere with an attendant sudden jump in the magnetic field. Such a jump propagates, presumably in the form of a hydromagnetic wave, from the outer magnetosphere to the ionosphere where it sets up a current system. The ionospheric current system induces a sharp charge in the earth's magnetic field and is visible almost simultaneously over the entire earth as a very sharp haze in the magnetogram. With modern instrumentation, using high-time resolution magnetometers, the SSC can be observed to propagate over the ground. In a coordinated data study, the world-wide response to such an event can be studied in great detail. The equivalent ionospheric current system can be mapped out, the nature of the "SSC" wave can be greatly clarified, and indeed, the "impulse responses" of the magnetosphere can be studied.

1.3 Programming background

From a programming point of view, this project involved two parts. Part One consisted of designing a data base suitable for this data set and the functional procedures to be performed, then inserting the data into the data base.

Part Two involved design and coding of the data analysis modules to be used during the conference.

1.4 Outline of Report

This report is divided into seven sections. Section 1 serves as an introduction and provides background material on the report. Section 2 describes the processing of the raw data into a data base suitable for the initiator's use at the conference. Section 3 contains a description of the data analysis modules created for the project, while Section 4 explains the preliminary tasks performed prior to the conference. The conference, itself, is described in Section 5, and the conclusions resulting from the conference are presented in Section 6. A final summary is given in Section 7.

2. DATA PROCESSING

2.1 Overview

This project involved the formation of the largest on-line Geomagnetic data base ever assembled for a specific conference. The size of the data base made it possible to formulate more general conclusions since a larger and more diverse sample set was available.

2.2 Data Description

The raw data tapes were sent to Bedford Research Associates from various participants. These tapes had a multitude of formats and

specifications. The data values were components of a vector, x, y and z from a station site for a specified time range.

Ten different participants from around the globe contributed this data; the number of station sites totaled 95. In order to increase the sample set, ten events were selected for study. Each station site contributed data for at least one event, while some provided data for all ten events.

2.3 Data Base Formation

Because the data base was to be used on-line, space was one of the major considerations in forming the data base. Ten separate data files were formed, since the ten events were totally unrelated and there would be no need to draw on data from different events at the same time.

The data bases were divided into two types of records: 1) header records containing pertinent information about the data, and 2) data records containing just packed data values. Constructing the data base in this manner enables one to search for information simply by scanning the header record. This reduces the running time needed for data analysis and gives faster results to the on-line user. Packing the data values in a suitable form greatly reduces the size of the data base.

For a detailed description of the data base structure, see APPENDIX I.

2.4 Problems Encountered

The main cause of problems stemmed from the diversity of data structures. Each of the ten participants supplied data in a different format, e.g. ASCII, EBCIDIC, character format, 16 or 32 bit binary, one component at a time, or 3 components at a time. Some participants used different formats for different events.

Each participant was asked to submit a sample data tape in advance so that Bedford Research might design, code and test a data conversion program for putting data into the data base. Participants were also asked to send listings or plots of the data for verification purposes. Some were unable to comply with these requests. A time table was formulated for preparation of the conference and a date was set for final submission of data by participants. Some data arrived late but was still processed in time for inclusion in the conference.

3. DATA ANALYSIS

3.1 Overview

In order to analyze the data in the most effective way, a group of statistical and data display procedures were selected. The ability to produce a listing of any data segment and also to graphically present this data were desirable features for direct interpretation of the data. Conference participants should also be able to do filtering, power spectra, cross spectra and hodograms of the data.

The first step involved production of data retrieval routines that would be time efficient. This accomplished, the design and coding of the actual analysis modules could be completed.

3.2 Analysis Programs

3.2.1 Printing the Data

The display of the data values in printed form was one of the most straightforward modules. By requesting a station site, date, start time, end time and data type, the specific data segment would be printed out. The available data types consisted of component x, y, z or the magnitude of any components taken - one, two or three at a time. (See Figure 3.1.)

3.2.2 Plotting the Data

The plotting module of the data analysis has the ability to produce standard plots or user-customized plots. The plots can be produced either on interactive graphics or on microfiche. The standard plots are multiple axis plots with axis scaling to fit the data, three plots per frame corresponding to the different components for an entire data file. (See Figure 3.2.)

To customize the plots, several options are available: station site, component or magnitude, and manual selection of axis scaling and base line. The final option is production of single axis plots. (See Figure 3.3.) This procedure can produce from 1-10 plots on a single axis.

3.2.3 Filtering Data

The filtering module was actually three-fold. The first part would filter the data to change the sampling rate. Because different participants used data at different sampling rates, it became necessary to design and construct a filter module that would create a uniform delta-time data base. Also, some of analysis required changing the sampling rate.

The second part was the construction of commonly used filters: low pass, high pass and band width. The filters would be used on a data set.

The third part was making a filter design program available to participants so they could design their own filters for use on a data set.

3.2.4 Power Spectra Analysis

A power spectra analysis module that would take the power spectra of a data set was put together. It was designed to print out a table of spectral characteristics and also produce custom plots with variable power and frequency axis ranges. (See Figures 3.4 and 3.5.)

3.2.5 Cross Spectra Analysis

A cross spectra analysis module with the ability to do power spectral estimates for up to 3 station sites was created. The module would also produce polarization parameters and cross spectral parameters.

3.2.6 Hodogram

See Figure 3.6

3.3 System Structure

A task-oriented system was constructed to access all the above modules. The system would perform the analysis module on the desired data file, access all the necessary system and data files, and produce the desired output files. With this system, the user could do an entire analysis module by issuing only one key-word command. This system also allowed the user to do a stream of analysis, i.e., concatenate analysis modules.

System modules were also added. One of these modules could construct a new data set by merging two old data sets. Another could change any information in the header record in a data set. For example, the user could change the station site name to one that was more descriptive. Finally, there was a module for listing out the station site information of all the data in the data bank. (Figure 3.7)

4. PREPARATION FOR THE WORKSHOP

4.1 Overview

Bedford Research Associates received the contract for this work at the AFGL conference in January 1981.

General specifications on the proposed analysis were received by the end of January. The system design was begun at that time, with the interface portions omitted until the data base structure could be determined.

By the middle of February, participants began sending in some sample data tapes. With these as a guideline, a first draft of the data base was constructed.

By mid-March the final data base structure was designed and a sample data base constructed. The printing and plotting analysis modules were completed and tested.

The months of March till the end of October were spent loading data into the data base and creating the other analysis modules.

4.2 Test Case Analysis

During the first two weeks in November, a final test case of the entire system was run. Every conceivable path of analysis was tested for complete accuracy. Also, every possible means of bringing down the system or "fooling" it was tried out in order to troubleshoot any problems that might occur during the conference.

4.3 Analysis Performed

The last two weeks in November were used for production runs that would be performed prior to the conference. This included plotting of all the data in the data base and forming a booklet of the plots.

For analysis purposes, it was decided to have all the data at a uniform sampling rate. Therefore, new data bases were created from the original by means of the filtering module.

Standard low pass, high pass and band width filters were also constructed for the conference.

5. WORKSHOP

5.1 Overview

The workshop ran quite smoothly, with language being the only major problem. Since participants came from around the world, communication was sometimes difficult.

5.2 Analysis Performed

Of the ten events available to participants, four were determined to be of special interest and were selected for the conference. All the analysis modules used were highly successful.

6. CONCLUSIONS

6.1 Programming Standpoint

From a programming standpoint, the conference was a total success. The participants were quite impressed and were pleased with the ability to perform a number of analysis techniques on such a diverse data base. The system ran perfectly and efficiently and was able to give the participants on-the-spot investigative tools for any and all suppositions brought up at the conference.

6.2 Technical Standpoint

From information gathered at the conference, participants were able to formulate outlines for a number of technical papers. Because of time limitations at the conference, final conclusions could not be drawn. However, participants will be interacting with the data base via mail and plan to reconvene in May to complete the writing of the technical papers.

Some of the participants have also submitted additional data that has since been added to the data base for further analytical studies.

7. SUMMARY

This conference was highly successful from all points of view. It gave the participants an opportunity to collectively form conclusions about the topic "Impulse Response of the Magnetosphere" and to test out these conclusions by on-line access to a large data base. It also provided the opportunity to gather information for personal as well as group suppositions. The conference served as a starting point for investigating suppositions and furnished a large data base which will be available for further investigation after the conference.

APPENDIX

I HEADER RECORD

Consists of 14 CDC integer words defined as follows:

1 = A10 station name

2 = I5 start data in form YYJJJ

3 = I5 end data in form YYJJJ

4 = I6 start time (UT) HHMMSS

5 = I6 end time (UT) HHMMSSS

6 = numerator of samples/second

7 = denominator of samples/second

8 = shift amount for X-component

9 = shift amount for Y component

10 = shift amount for Z-component

11 = scaling factor for X-component

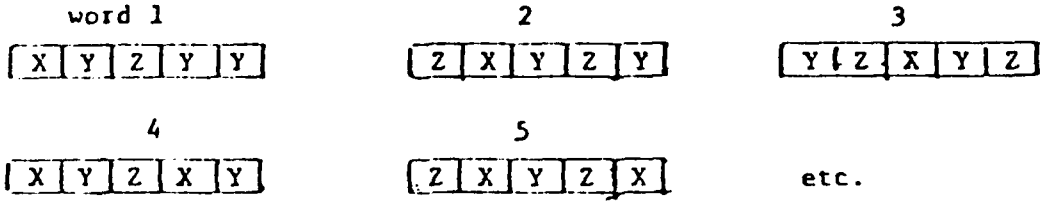
12 = scaling factor for Y-component

13 = scaling factor for Z-component

14 = default value for this station

II DATA RECORD

Consists of 384 CDC words packed five values/word in the following format:



III CONVERT FROM REAL VALUES TO UNPACKED DATA VALUES

a) Final MID Value = (max + min)/2

b) Determine range of values
RANGE = max - min.

c) By using a scale factor, adjust range to be less than 4,000.

d) Do a, b, c above for all 3 components, then form HEADER RECORD where

HEAD (11) = - Scale Factor for X
HEAD (12) = - Scale Factor for Y
HEAD (13) = - Scale Factor for Z

If scaling operation to adjust the range is division, then

HEAD (11) = Scale Factor for X
HEAD (12) = Scale Factor for Y
HEAD (13) = Scale Factor for Z

e) Buffer out Header (1) - Header (14) to tape 2.

f) Form integer data array of adjusted values by
adjusted = [original - MID Value]
in the multiplication scale
divisional scale
in the form of triplets for each Dt.

DATA (1) = X (t_i = start + iDt)
DATA (2) = Y ½ t_o
DATA (3) = Z
DATA (4) = X
DATA (5) = Y ½ t_i
DATA (6) = Z

g) Where there are 1920 data values in the array (i.e., 640 values of X,Y,Z respectively), call subroutine FILL.

The calling sequence is
Call Fill (DATA,NUM,TOTAL,HEADER)

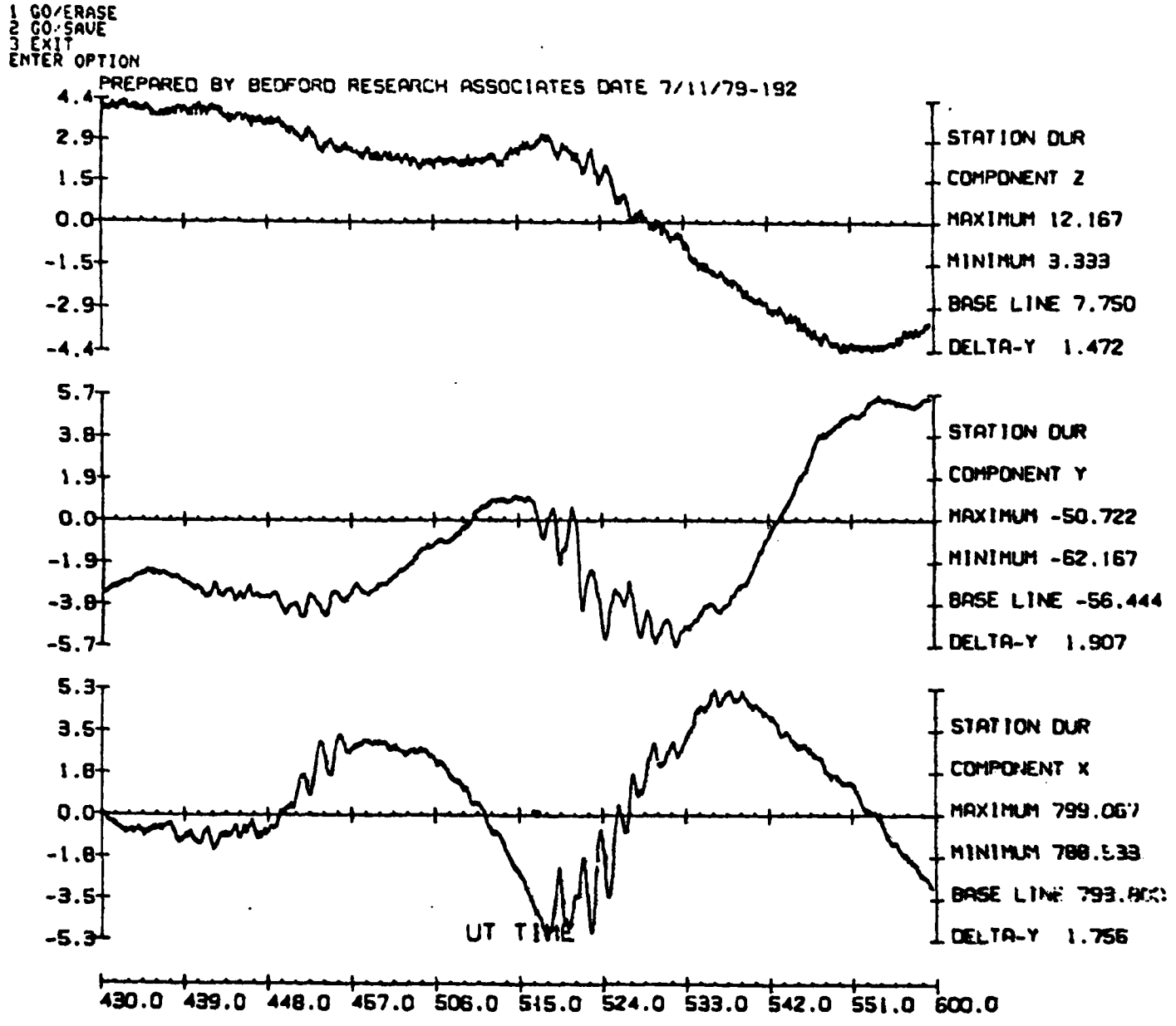
DATA = 1920 or less integer values.
NUM = number of data values in array data (Integer)
ITOTAL = returned value of number of data records buffered out. ITOTAL must be originally set to zero and is incremented by 1 for each call to fill.
HEADER = default value

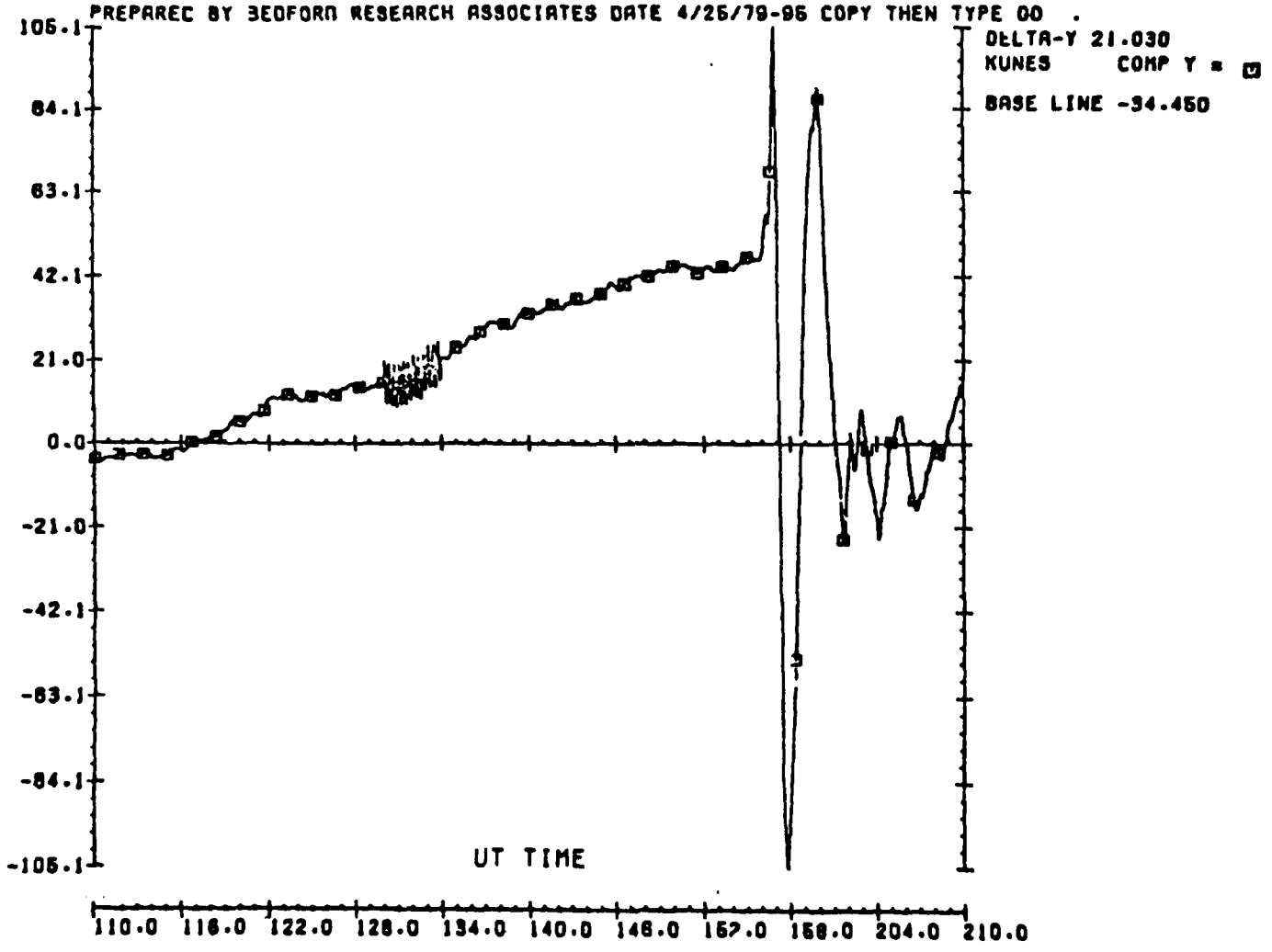
h) If your last batch of data does not fill up the data array of 1920 values, subroutine FILL fills in all values between NUM and 1920 with HEAD (14) before packing and buffering out the data.

STATION KEV COMPONENT X DATE 03/08/78-067 START TIME 150000 END TIME 150500

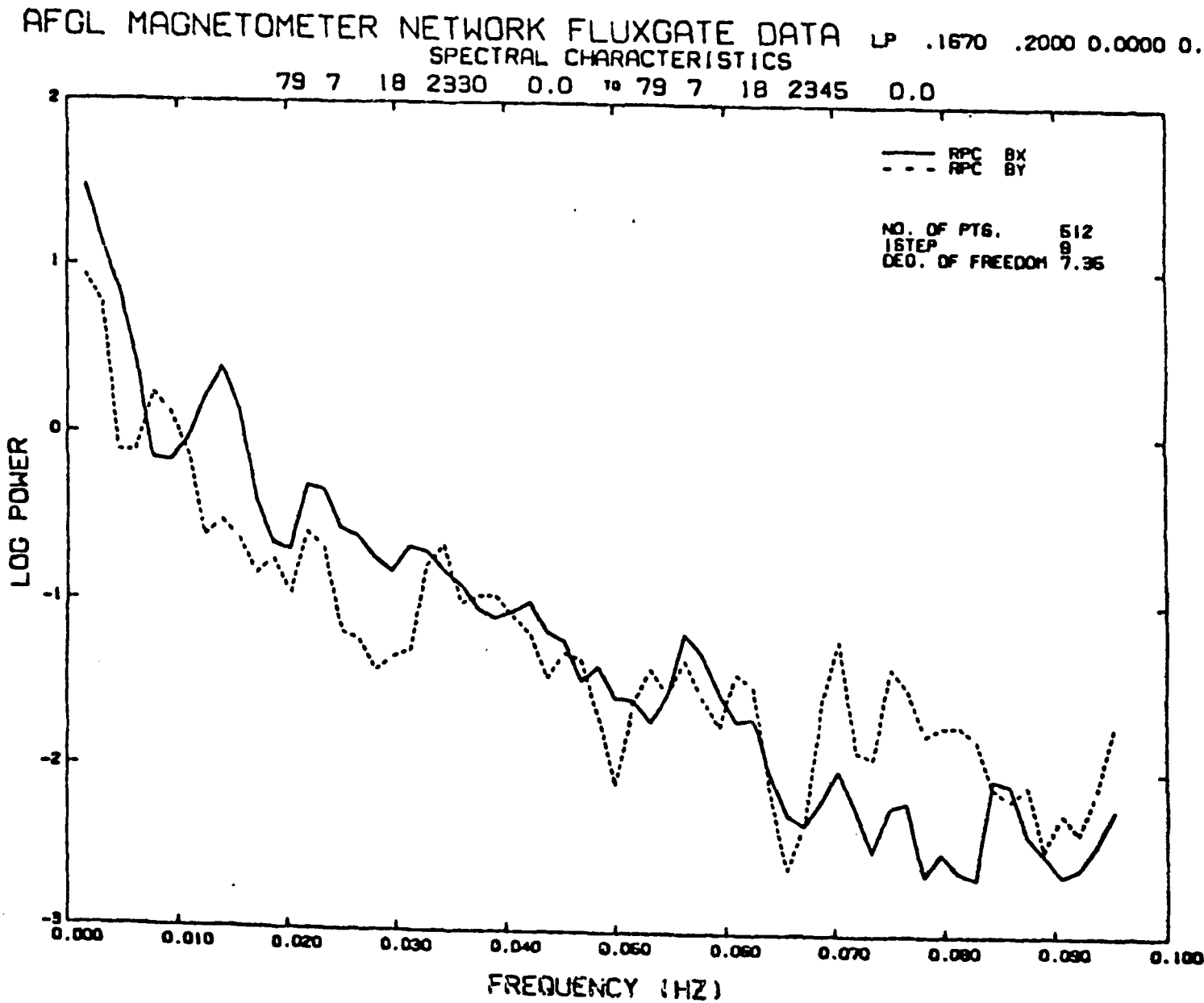
SAMPLES PER SECOND 1.000

TIME	DATA	DATA	DATA	DATA	DATA	DATA	DATA	DATA	DATA	DATA
150000	1.4	1.4	1.4	1.4	1.5	1.5	1.6	1.6	1.7	1.8
150010	1.8	1.9	2.0	2.2	2.4	2.5	2.5	2.6	2.7	2.6
150020	2.5	2.4	2.3	2.2	2.2	2.1	2.1	2.1	2.1	2.1
150030	2.1	2.1	2.1	2.0	1.9	1.8	1.7	1.7	1.6	1.5
150040	1.5	1.5	1.4	1.4	1.4	1.3	1.3	1.3	1.2	1.1
150050	1.0	.8	.8	.7	.5	.4	.3	.1	.1	0.0
150100	0.0	-.1	-.1	-.1	-.2	-.2	-.4	-.5	-.6	-.7
150110	-.9	-1.0	-1.1	-1.2	-1.3	-1.4	-1.6	-1.7	-1.9	-2.1
150120	-2.3	-2.4	-2.5	-2.7	-2.8	-2.9	-2.9	-3.0	-3.1	-3.0
150130	-3.0	-3.0	-3.0	-3.0	-2.9	-2.9	-2.8	-2.7	-2.6	-2.6
150140	-2.5	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.3	-2.2
150150	-2.2	-2.0	-1.9	-1.8	-1.6	-1.5	-1.3	-1.1	-.9	-.8
150200	-.6	-.4	-.3	-.1	.1	.2	.4	.6	.8	1.0
150210	1.3	1.5	1.7	1.9	2.1	2.2	2.4	2.5	2.6	2.7
150220	2.7	2.8	2.8	2.9	3.1	3.2	3.3	3.4	3.6	3.7
150230	3.8	4.0	4.1	4.3	4.4	4.6	4.6	4.6	4.7	4.6
150240	4.7	4.7	4.8	4.9	4.9	4.9	5.0	5.0	4.9	4.8
150250	4.8	4.8	4.7	4.8	4.8	4.8	4.6	4.5	4.3	4.1
150300	3.9	3.7	3.5	3.4	3.2	3.1	3.0	2.8	2.7	2.6
150310	2.4	2.3	2.2	2.1	2.0	1.9	1.8	1.8	1.8	1.7
150320	1.7	1.6	1.5	1.4	1.3	1.3	1.3	1.3	1.3	1.2
150330	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.2	1.1	1.1
150340	1.1	1.1	1.1	1.2	1.2	1.3	1.4	1.4	1.5	1.6
150350	1.7	1.8	1.9	1.9	2.0	2.1	2.2	2.3	2.3	2.4
150400	2.5	2.5	2.6	2.6	2.8	2.6	2.7	2.8	2.8	2.9
150410	3.0	3.1	3.1	3.1	3.2	3.3	3.4	3.4	3.5	3.6
150420	3.6	3.7	3.8	3.8	3.9	4.0	4.0	4.0	4.2	4.3
150430	4.4	4.5	4.7	4.8	4.9	4.9	5.1	5.2	5.3	5.4
150440	5.4	5.5	5.6	5.7	5.8	5.8	5.9	6.0	6.0	6.1
150450	6.2	6.3	6.4	6.5	6.5	6.6	6.7	6.7	6.9	6.9

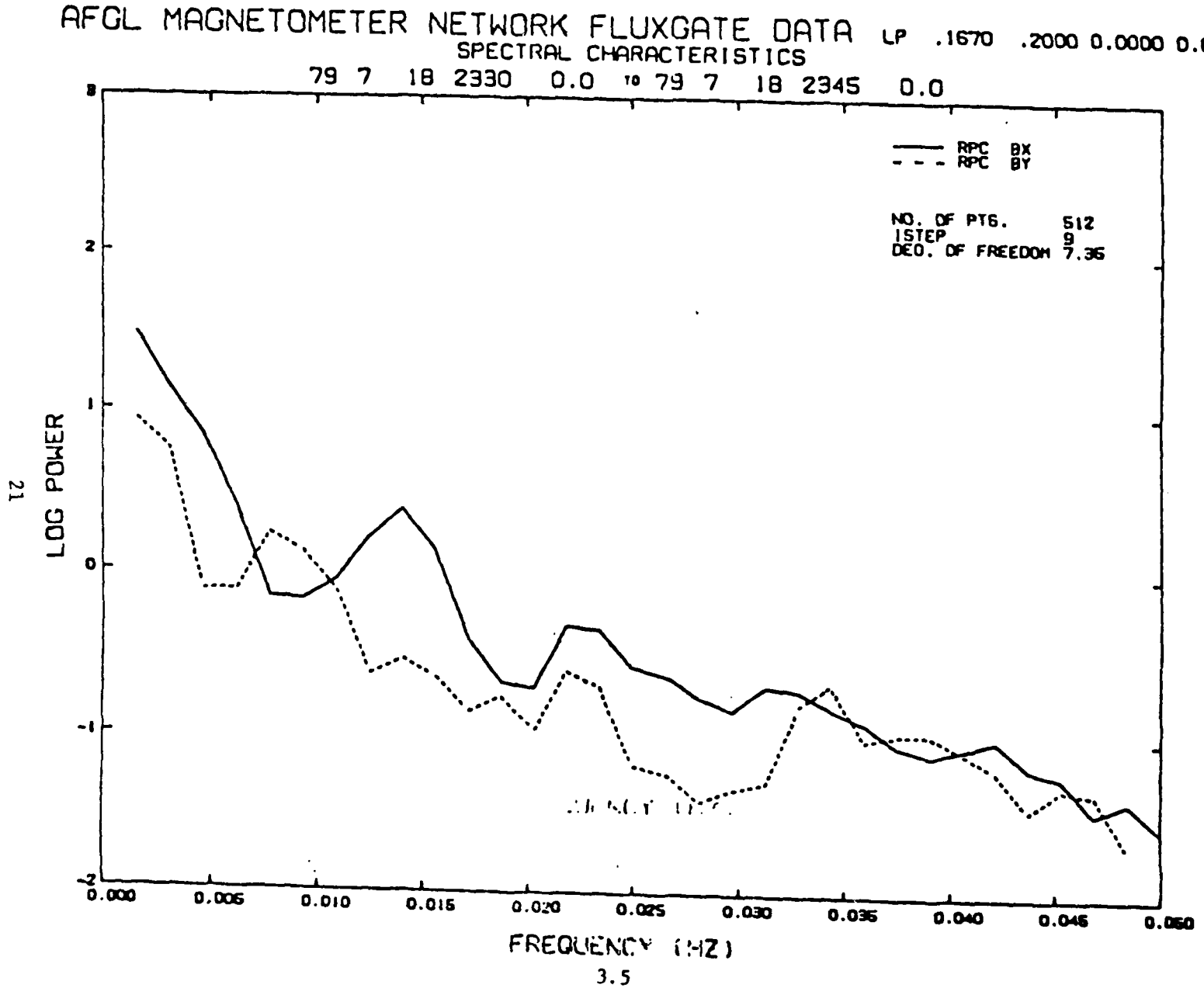

1 GO/ERASE
2 GO-SAVE
3 EXIT
ENTER OPTION
PREPARED BY BEDFORD RESEARCH ASSOCIATES DATE 7/11/79-192
4.4
2.9
1.5
0.0
-1.5
-2.9
-4.4
STATION DUR
COMPONENT Z
MAXIMUM 12.167
MINIMUM 3.333
BASE LINE 7.750
DELTA-Y 1.472
5.7
3.8
1.9
0.0
-1.9
-3.8
-5.7
STATION DUR
COMPONENT Y
MAXIMUM -50.722
MINIMUM -62.167
BASE LINE -56.444
DELTA-Y 1.907
5.3
3.5
1.8
0.0
-1.8
-3.5
-5.3
STATION DUR
COMPONENT X
MAXIMUM 799.067
MINIMUM 788.533
BASE LINE 793.800
DELTA-Y 1.756
UT TIME
430.0 439.0 448.0 457.0 506.0 515.0 524.0 533.0 542.0 551.0 600.0


PREPAREC BY 3EOFORD RESEARCH ASSOCIATES DATE 4/26/79-96 COPY THEN TYPE 00 .
DELTA-Y 21.030
KUNES COMP Y =
BASE LINE -34.450
105.1
84.1
63.1
42.1
21.0
0.0
-21.0
-42.1
-63.1
-84.1
-105.1
UT TIME
110.0 116.0 122.0 128.0 134.0 140.0 146.0 152.0 158.0 204.0 210.0

3.3


AFGL MAGNETOMETER NETWORK FLUXGATE DATA LP .1670 .2000 0.0000 0.0000 6 6
SPECTRAL CHARACTERISTICS
79 7 18 2330 0.0 TO 79 7 18 2345 0.0
RPC BX
RPC BY
NO. OF PTS. 512
ISTEP 9
DEG. OF FREEDOM 7.36
LOG POWER
FREQUENCY (HZ)
2
1
0
-1
-2
-3
0.000
0.010
0.020
0.030
0.040
0.050
0.060
0.070
0.080
0.090
0.100

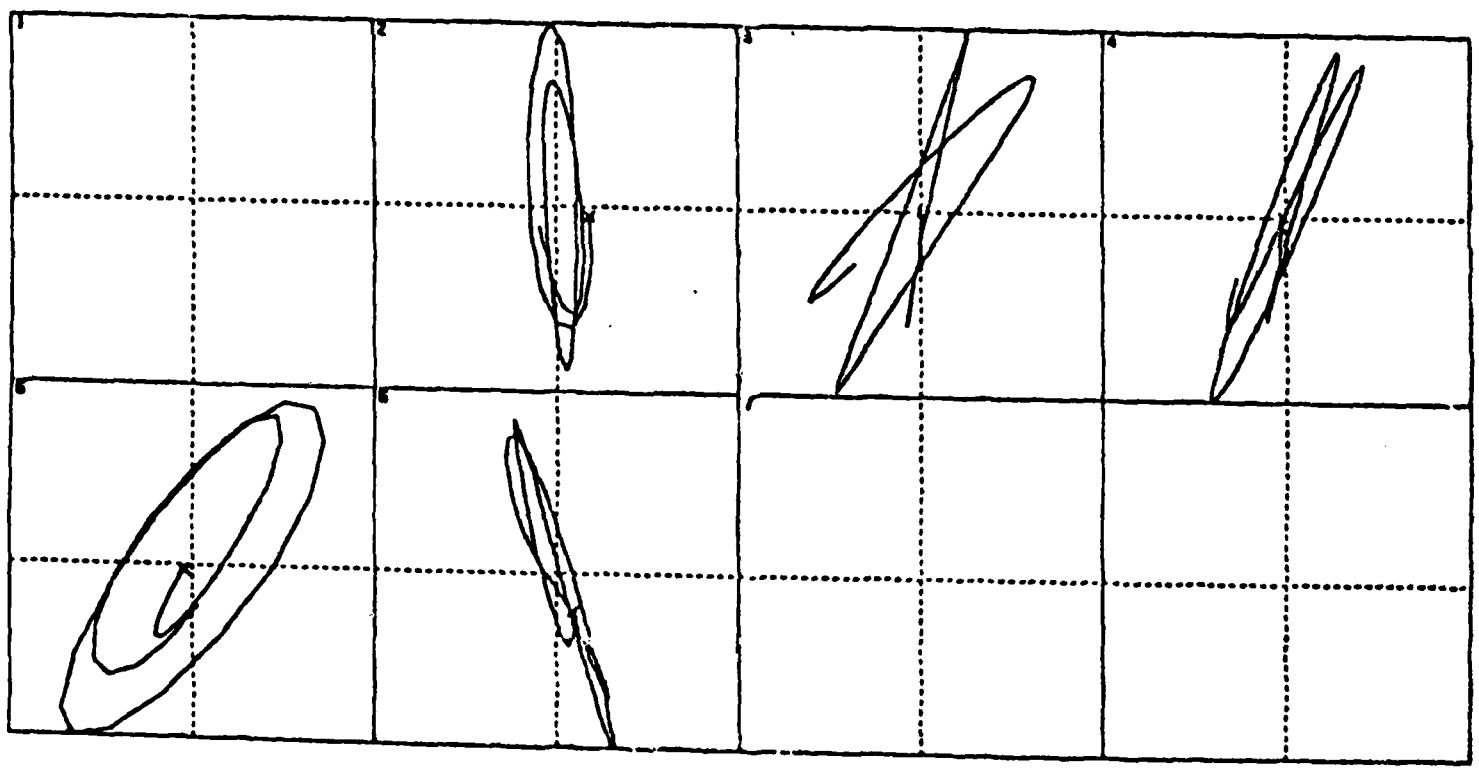
3.4

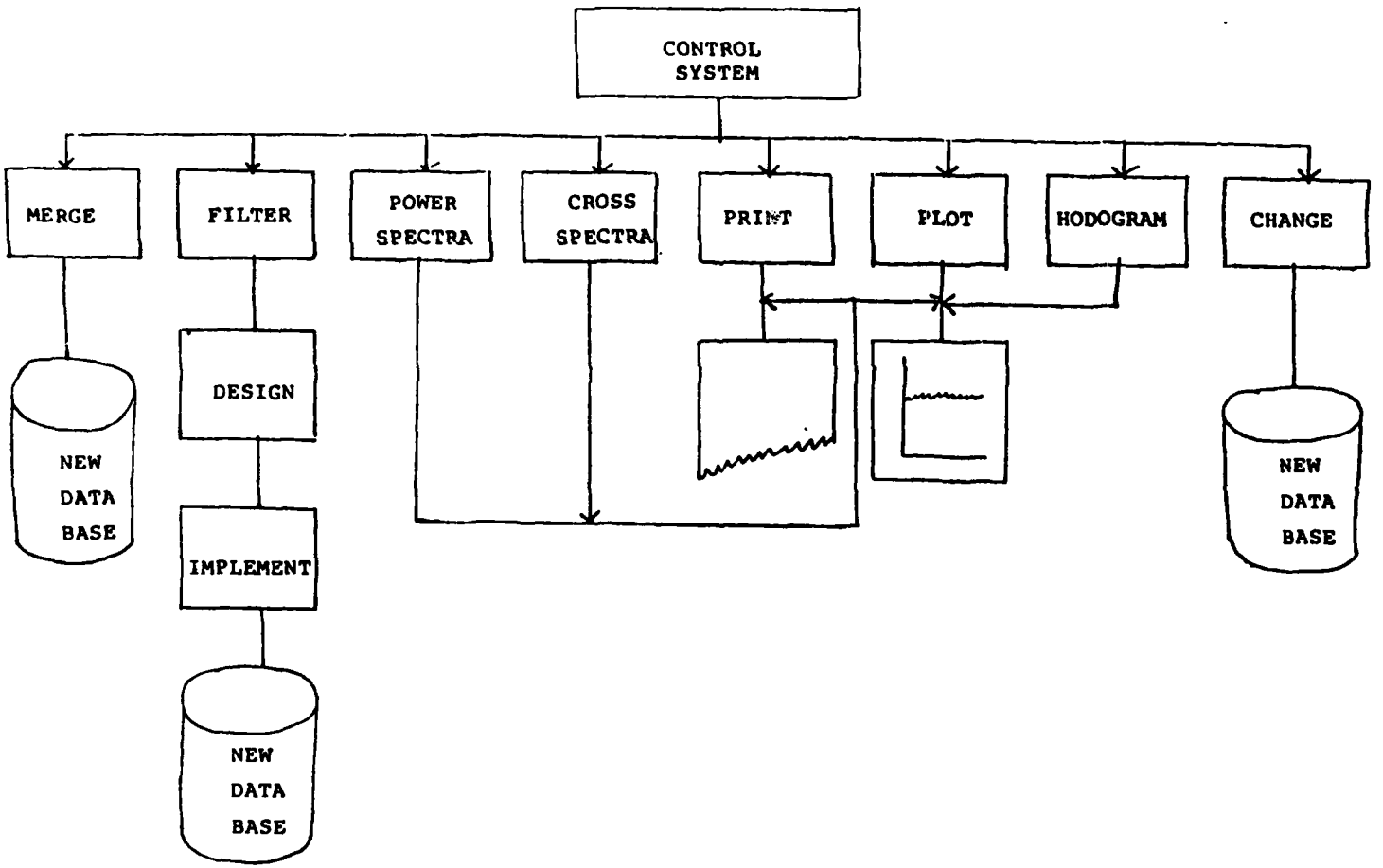

AFGL MAGNETOMETER NETWORK FLUXGATE DATA LP .1670 .2000 0.0000 0.0000 5 5..
SPECTRAL CHARACTERISTICS
79 7 18 2330 0.0 TO 79 7 18 2345 0.0
RPC BX
RPC BY
NO. OF PTS. 512
ISTEP 9
DEG. OF FREEDOM 7.36
LOG POWER
3
2
1
0
-1
-2
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
0.045
0.050
FREQUENCY (HZ)

AFGL MAGNETOMETER NETWORK FLUXGATE DATA BP .0080 .0100 .0200 .0220 1 6.0

79 7 18 2332 0.0 to 79 7 18 2335 0.0

1.	-		SCALE = GAMMA
2.	RPC BY	RPC BX	SCALE = 0.89 GAMMA
3.	CDS BY	CDS BX	SCALE = 1.06 GAMMA
4.	MCL BY	MCL BX	SCALE = 1.74 GAMMA
5.	SUB BY	SUB BX	SCALE = 1.49 GAMMA
6.	LOC BY	LOC BX	SCALE = 0.61 GAMMA
7.	-		SCALE = GAMMA




CONTROL SYSTEM
MERGE
FILTER
POWER SPECTRA
CROSS SPECTRA
PRINT
PLOT
HODOGRAM
CHANGE
NEW DATA BASE
DESIGN
IMPLEMENT
NEW DATA BASE
NEW DATA BASE

MINUTEMAN DATA ANALYSIS

1. INTRODUCTION

The Minuteman database consists of recordings taken from 20 different sites. These sites are labelled I02-111 and R02-111. The time span for all sites where data was recorded is December 1, 1979 to March 31, 1980.

Each site has ten readings or variables. These variables consist of X, Y and Z platform, instrument, and gyro, plus a phi-N measurement. Each variable has a sampling rate of seven hours and fifteen minutes (± 30 minutes). The sampling times for all ten variables of one site are the same, but each site has different sampling times. Consequently, each site has gaps in the data unique to that particular site.

2. OBJECTIVES

The current analysis of the Minuteman database was directed toward locating consistent structures in the data. All analysis was focused on relationships between variables of the same site that were exhibited in a large proportion of the twenty sites. Direct relationships between variables of different sites were not studied. For example, a coherence function using X-platform data from site I02 and I03 as input was not performed.

There was no effort to detect or analyze events in the data. An event is defined as one or more large impulses or oscillations relative to the majority of data values in a continuous "stream" of data. (Figure 1A depicts an "event" occurring from days 73 to 80).

It was desired to eliminate any redundant variables by observing consistently high correlation values among variables in a site. A multiple step-wise regression holding phi-N dependent will be implemented to attain this goal. Power spectra and coherence functions attained by using an FFT algorithm should detect any strong frequencies relating to the variables.

3. ANALYSIS

The first effort consisted of plotting the ten variables of site IO2. This site was chosen at random to get some idea of the data structure. The plots revealed that each variable consists of roughly evenly-spaced stretches (i.e., no gaps) of data with an average of fifteen points or a spanning of approximately five days. These "stretches" were interrupted by gaps of approximately one day, but occasionally as long as twenty days. Also, each stretch of evenly-spaced samples appeared to have its own bias associated with it. A check of the output listing on all twenty sites confirmed this trait.

A multiple regression algorithm was implemented at this time. The nine variables (X, Y, and Z platform, instrument and gyro) were the independents, while phi-N was held as the dependent variable. Each evenly-spaced stretch of observations for all sites and variables had its associated bias removed from it. The results indicated an extremely strong correlation between X-platform and the X, Y, and Z instrument measurements. When the variables were not forced into the regression model, one of the four variables mentioned above was usually entered into the model. The squared multiple correlation coefficient (R2) was usually .05 for the twenty sites with a maximum of .15.

The regression package was applied to all twenty sites forcing the nine variables into the model. The results were not significantly different from those with no forcing.

The tables depicting the regression results were drawn. Table 1A illustrates which variables were entered into the unforced step-wise model and the order in which they were inserted for each site. Also, the squared multiple correlation coefficient for each site is listed. Table 1B has a similar setup (sites versus variables), except that it indicates pairs of variables whose common correlation coefficient is greater than 0.4. Table 1C contains counts of two variables with a correlation coefficient larger than 0.4 out of the 20 sites.

Upon inspection of the time gaps of all sites, it was observed that a large gap exists between December 19 and January 15 for all sites. All subsequent analysis was performed only on data occurring after this gap since no similar gaps exist from January 15 to March 31. The data base was further reduced by eliminating the three instrument variables. This was done because of the large and consistent correlation of these variables to the X-platform.

Two approaches were decided upon at this time: first, regression analysis on the reduced database with forcing only on the X-platform, and second, spectra analysis and coherence functions on the X-platform and the phi-N variable. These two variables are linked by stronger correlation values between the X-platform and instrument readings to the phi-N measurements relative to the remaining variables.

The results of the regression analysis on the reduced database is depicted in Tables 2A, 2B, and 2C. Their structure is similar to Tables 1A, 1B, and 1C, respectively. The results revealed no consistent correlations between any of the seven remaining variables. No increase in the squared correlation coefficient was observed.

The reason for the spectra analysis was to check whether the signals have low frequency peaks. If this is the case, then a low-pass filter can be implemented to reduce higher frequency noise. Regression can be performed on the filtered data. Hopefully, this will yield better correlation and R2 results.

Before spectra analysis (i.e., FFT) can be applied to the data, the signal samples must be evenly spaced in time. The Minuteman data must have all gaps interpolated before an FFT can be applied. A linear prediction routine was used to interpolate for missing data. Specifically, forward and backward prediction was used with weights determined by the estimated variance associated with the measured data points.

The gaps in all X-platform and phi-N signals were filled. The coherence and power spectrum plots were then obtained. The results of the plots displayed only one prominent characteristic. A large "spike" was evident in the spectra of the X-platform at a frequency of one cycle per 21.75 hours.

The spike was observed in all X-platform plots but was not consistent in either the coherence plots or the phi-N plots. This feature was consistent with a parameter associated with sampling times. Each time data was recorded a "direction" (east or west) associated with that observation. Every third observation was associated with an east orientation. Upon close inspection of the data, especially the X and Y platforms, it was noted that different bias levels existed for each direction. Since the sampling rate was 7.25 hours and the bias occurred in cycles of three observations, a frequency of one cycle per 21.75 hours should have been and indeed was prominent in the spectra.

An algorithm was designed to remove this east-west bias. The data for each stretch of evenly-spaced data was grouped by its direction, and the mean for each group was calculated and subsequently removed. This was performed on all variables for all sites and regression analysis was reapplied. The results showed some improvement in the correlation values between the X and Y platform and the dependent variable phi-N. No improvement was observed in the R2 values. Power spectra and coherence plots on this database illustrated the attenuation of the frequency component due to the east-west biasing, but no other frequencies were significantly stronger on a consistent basis.

Since the coherence function has a range of 0 to +1, all coherence plots could be averaged together. Coherence averaging plots were generated for cases with bias present as well as removed. The results from both plots revealed no significant frequencies common to the X-platform and phi-N measurements.

4. CONCLUSIONS

Thirty percent of the database can be reduced by eliminating the three instrument variables. This is true because of the consistent high correlation between the X-platform variable and the three instruments. The database can be further reduced by ignoring samples prior to January 15. Any subsequent analysis should take into account these features.

No significant relationships were exhibited between the remaining independent variables and phi-N. It should be noted that only structures common to all twenty sites were sought.

The power spectra indicated spikes randomly dispersed through the frequency spectrum. This ruled out any digital filtering of the data to improve correlation values; real spikes could not be distinguished from noise. The spectra indicated different bias levels only in association with the east-west orientations.

It should be repeated that all studies were made on relationships between variables of the same site that could be validly extended to all sites. There was no detection or analysis of events occurring in variables.

INDEPENDENTS

No forcing results

SITE	X-PLATFORM	Y-PLATFORM	Z-PLATFORM	X-INSTRU	Y-INSTRU	Z-INSTRU	X-GYRO	Y-GYRO	Z-GYRO	PHI-N	R^2
I02								1			.020
I03						1		2	3		.080
I04						1					.045
I05				1							.018
I06											.000
I07											.000
I08			2		1			3		Dependent Variable	.150
I09	1										.029
I10	1										.034
I11					1						.028
R02				1							.052
R03								1	2		.059
R04											.000
R05		2						1			.067
R06				2		3	1				.067
R07	1										.026
R08			1								.030
R09											.000
R10						1			2		.083
R11								1			.032

TABLE IA

C.C ≥ .4

SITE	1 X-PLATFORM	2 Y-PLATFORM	3 Z-PLATFORM	4 X-INST	5 Y-INST	6 Z-INST	7 X-GYRO	8 Y-GYRO	9 Z-GYRO	10 PHI-M
I02	4,5,6			1,5,6	1,4,6	1,4,5				
I03	4,5,6	5	6	1,6	1,2	1,3,4				
I04	4,6	5,9		1,6	2	1,4				
I05	2,4,5,9	1,4,5,9		1,2,5,9	1,2,4,9		8	7	1,2,4,5	
I06	4,5,6	5		1,5,6	1,2,4	1,4				
I07	4,5,6	5		1,5,6	1,2,4,9	1,4			5	
I08	4,5,6	5		1,5,6	1,2,4,6	1,4,5				
I09	4,5,6	4,5		1,2,5,6	1,2,4	1,4		9	8	
I10	4,5,6	5	5,6	1,5,6	1,2,4,6	1,3,4,5	9		7	
I11	4,5,6	5	5,6	1	1,2,3	1,3,9	9	9	6,7,8	
I12										
R02	3,4,5,6		1,5,6	1,5,6	1,3,4,6	1,3,4,5	8	7		
R03	4,5,6	5	6	1,5,6	1,2,4,6	1,3,4,5				
R04	4,5,6	5		1,5,6	1,2,4,9	1,4			5	
R05	4,5,6	5		1	1,2	1		9	8	
R06	3,4,5,6		1,6	1,5	1,4	1,3	9		7	
R07	4,5,6	3,5	2,5	1,6	1,2,3	1,4				
R08	4,5,6,8			1,5,6,8	1,4,6	1,4,5,8		1,4,6		
R09	2,4,5	1,4,5	3	1,2,5	1,2,4	3	8	7		
R10	4,5,6		6	1,5,6	1,4	1,3,4				
R11	4,5,6	5,8	6	1,5	1,2,4	1,3		2,9	8	

TABLE 1B

C.C $\geq$ .4

VARIABLES	X-PLATFORM	Y-PLATFORM	Z-PLATFORM	X-INST	Y-INST	Z-INST	X-GYRO	Y-GYRO	Z-GYRO	PHI-N
	1	2	3	4	5	6	7	8	9	10
1	-	2	2	20	19	18	0	1	1	0
	1	2	3	4	5	6	7	8	9	10
2	2	-	1	3	14	0	0	1	2	0
3	2	1	-	0	4	8	0	0	0	0
4	20	3	0	-	15	14	0	1	1	0
5	19	14	4	15	-	6	0	0	3	0
6	18	0	8	14	6	-	0	1	1	0
7	0	0	0	0	0	0	-	3	3	0
8	1	1	0	1	0	1	3	-	4	0
9	1	2	0	1	3	1	3	4	-	0
10	0	0	0	0	0	0	0	0	0	-

TABLE 1C

SITE		INDEPENDENT X-PLAT	Y-PLAT	Z-PLAT	X-GYRO	Y-GYRO	Z-GYRO	PHI-N	R^2
I02	F								.004
I03	F								.054
I04	F								.037
I05	F								.008
I06	F								.017
I07	F								.004
I08	F	2							.147
I09	F								.038
I10	F								.026
I11	F								.013
R02	F								.053
R03	F								.038
R04	F								.020
R05	F	2				3			.121
R06	F				2				.048
R07	F								.021
R08	F		2						.052
R09	F	2							.073
R10	F								.041
R11	F					2			.033

Dependent Variable

TABLE 2A

C.C. $\geq .4$

SITE	X-PLAT	Y-PLAT	Z-PLAT	X-GYRO	Y-GYRO	Z-GYRO	PHI-N			
	1	2	3	4	5	6	7			
I02	4			1	6	5				
I03										
I04		6		5	4	2				
I05	2,6	1,6				1,2				
I06										
I07	6	6				1,2				
I08										
I09	2	1			6	5				
I10	2	1		6		4				
I11	2	1		6	6	4,5				
R02	3		1	5	4					
R03										
R04	6	6				1,2				
R05					6	5				
R06	3		1	6		4				
R07										
R08	5,6				1	1				
R09				5	4					
R10										
R11		5			2,6	5				

TABLE 2B

C.C. $\geq$ .4

SITE	X-PLAT	Y-PLAT	Z-PLAT	X-GYRO	Y-GYRO	Z-GYRO	PHI-N
1	-	4	2	1	1	4	0
2	4	-	1	0	0	4	0
3	2	1	-	0	0	0	0
4	1	0	0	-	3	3	0
5	1	0	0	3	-	5	0
6	4	4	0	3	5	-	0
7	0	0	0	0	0	0	-

TABLE 2C

TERRAIN EFFECTS ON TURBULENCE MODEL

1. INTRODUCTION

An attempt is made to develop regression terms and quantify the effects of mountain-induced disturbances in the lower atmosphere. This information will be used to improve an existing model for predicting the occurrences of turbulences in the lower atmosphere. Murphy and Scharr (1981) and Murphy et al (1982) contain further explanations of the technique for determining turbulence estimates and the present model for predicting occurrences.

The gradient Richardson number, used as the dependent variable in the regression analysis, is:

$$Ri = \frac{(g/T)\ [dT/dZ + \Gamma_d\ ]}{(\ \partial V_x/\ \partial Z)^2\ + (\ \partial V_y\ /\ \partial Z)^2}$$

where g is the acceleration of gravity, T is the absolute temperature, dT/dZ is the vertical temperature gradient, Γ_d is the dry adiabatic lapse rate (9.7° K/Km) and the denominator is the square of the vertical shear of the horizontal wind. Rawinsonde measurements of wind and temperature as a function of altitude are used to determine Richardson numbers. The Richardson number values as a function of altitude are calculated at 1 km levels from 2 km to 30 km for each of two daily height profiles of wind and temperature. This provides from 100 to approximately 180 values of Ri for each altitude bin per season. The percent occurrence of turbulence is obtained by taking the ratio of the number of occurrences of $Ri \leq 1.0$ to the total number of measurements per season. The likelihood of occurrence of turbulence is then based on the relative frequency of occurrence of a critical Richardson number ($Ri_c = 1.0$).

The present model for predicting turbulence has been established for the following ranges: lower range (2-7 km), middle range (8-13 km) and upper range (14-19km). Model variables include location type (latitude, longitude, altitude), transformations of them (lat^i, $long^j$, alt^k); i,j=0,1,2; k=0,1,2,3), cross products terms (lat^i, $long^j$,alt^k) and seasonal indicators. The dependent variable, an indicator of turbulence, is based on percent occurrences of the Richardson number less than unity. By using location and seasonal information we have been able to explain a substantial amount of the variation in the turbulence indicator (38% in the lower range, 61% in the middle range, and 42% in the upper range). This has been done for a sample 21 stations chosen as representative of the continental United States. Also, another important result of that study is that yearly variations in the occurrences of Ri_c are found to be small (less than 1% of the total variation). Evidence of mountain-induced disturbances are presented and improvements are made in the model through the use of terms to quantify terrain aspects.

2. EVIDENCE OF MOUNTAIN-INDUCED DISTURBANCES - A COMPARISON OF DATA SETS

A qualitative analysis performed on many of the data reporting stations (81 in the continental United States) revealed that the height profiles of percent occurrences of $Ri \leq 1.0$ for stations in the proximity of mountain ranges are markedly different from those stations in the plains areas of the United States. This was noted from visual observation of microfiche computer plots.

Three sets of data are used to describe mountain-induced disturbances in the atmosphere. Each set consists of 3 stations at nearly the same latitude. In each set, two measuring stations are in the midwestern plains and one is in the western mountain region. The one exception to this is Chatham, Massachusetts, which is here considered a plains station. Only nine stations from the 81 stations in the continental United States are suitably located to provide this kind of comparison. In the first set, Figure 1, the three stations used for comparison are within approximately 2 degrees of latitude. The two plains stations, Sault St. Marie, Michigan

and International Falls, Minnesota are plotted together in Figure 1a. The degree of closeness of fit for the two profiles of percent occurrence of $Ri \leq 1$ are evident. This is true for all four seasons. Note that the profiles for each season are for the five-year averages (1971-1975). Each of the years compared separately is also close. This is true since, as was previously stated, the yearly contribution to variation in percent occurrence of turbulence (using Ri_c as an indicator) was found by regression to be less than one percent. The higher percent occurrences for the mountain site, Great Falls, Montana, shown in Figure 1b, continues up to approximately 15 km. In fact, percent occurrences are consistenly higher in all of the five years from 1971-1975.

The two plains stations of the second set, Topeka, Kansas and Dayton, Ohio, are plotted together in Figure 2a. The mountain station for this set is Denver, Colorado, which is plotted for comparison along with Topeka, Kansas in Figure 2b. Here again there is fairly good agreement between the two plains stations, whereas there are differences as high as 25% between the plains and mountain stations. These differences are found as high as 8 km. The three stations in this set are located within a degree of latitude.

In the third set, the three stations are separated by just over one degree in latitude. The plains stations, Chatham, Massachusetts and Flint, Michigan, are plotted together in Figure 3a and are remarkably close. In comparison, the percent occurrences of the mountain station, North Platte, Nebraska (Figure 3b), are seen to be markedly higher than the plains station, Flint, Michigan. This is true up to a height of 5 km.

3. TERRAIN EFFECTS ANALYSIS - STATISTICAL APPROACH

Appropriate statistical techniques were employed to determine significance and to develop the improved model. Using the 21 stations that determined the original model, the following points are considered in an effort to quantify the terrain effects:

(a) Classification of terrain features
(b) Creation of terrain type variables
(c) Determination of effects

(a) Classification

It is apparent from a topographic map of the United States that there are three types of terrain - the coastal sea level areas, the central plains section and the western mountain region. Initial classification was done by grouping stations within the eastern, central, and western section. To determine if terrain effects were significant, a two-way analysis of variance (ANOVA) test was performed for each altitude bin (Draper and Smith, 1981). The range of the mean differences in turbulence within each altitude bin for the different terrain groups was of interest. This statistical test considers mean differences between groups using the test statistic F. Results of the two-way ANOVA on the 2-7 km and 8-13 km bins are presented (Table 1). Examination of the F-statistic reveals that there are overall significant differences ($p<<.001$) between the defined terrain groups (Table 2).

(b) Creation of terrain variables

In order to include the effects of terrain in the model, two variables were created to indicate the type of terrain:

$$\text{sea level} \begin{cases} 1 & \text{if station is in sea level category} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{mountain} \begin{cases} 1 & \text{if station is in mountain category} \\ 0 & \text{otherwise} \end{cases}$$

A plains station would then be indicated by sea level = 0 and mountain = 0. A multiple stepwise least squares regression procedure was

employed to determine the significance of terrain differences on turbulence. The set of independent variables candidates for entry into the turbulence model include the previous terms from the existing model plus sea level and mountain terms. Based on the ANOVA results, the regions 2-7 km and 8-13 km were selected for modeling.

(c) Determination of effects

In the 2-7 km range, the multiple correlation coefficient squared (R^2) increased by nearly 10% over the model without terrain indicators. Thus, a significant improvement in the turbulence model is represented.

There was a negligible contribution to turbulence explained in the 8-13 km model using the terrain indicator variables. This was true even though the ANOVA test in that range determined significant differences between terrain groups. The explanation for this is that turbulence variations due to differing terrain types occur at varying altitudes in the 8-13 km range, thus accounting for the ANOVA result. Because there is no sustained effect at any given altitude, the terrain variables do not add significant information to explain turbulence in the 8-13 km bins. Occurrences of turbulence estimates, based on a critical Richardson number of 1.0, are shown in comparison to model predicted values in Figure 4. These results from three representative stations, two in the mountain and one in the plains region, clearly demonstrate the improvements in the model with the added terrain terms.

4. SUMMARY

a. An independent variable, quantifying terrain characteristics in the general region of station location, improves turbulence estimates in the 2-7 km altitude range. The accuracy of the estimates is increased by 10% over the model without terrain indicators.

b. Turbulence estimate above 8 km altitude are not significantly improved by knowing the type of terrain.

c. There is evidence to suggest that a more localized effect is present. The average heights of the mountains within a radius of several hundreds of km of a station appear, from a qualitative inspection, to be related to the degree of activity in the 2-7 km altitude region. This, however, is difficult to quantify and further efforts will be applied in this direction.

TABLE 1

TWO-WAY ANALYSIS OF VARIANCE RESULTS

2-7 km	df	Winter	Spring	Summer	Fall
terrain	2,108	10.91 (7.41)*	4.99 (7.41)	28.28 (7.41)	24.54 (7.41)
alt. level	5,108	4.78 (4.48)	9.87 (4.48)	5.50 (4.48)	4.77 (4.48)
8-13 km					
terrain	2,108	30.42 (7.41)	36.78 (7.41)	6.72 (7.41)	15.38 (7.41)
alt. level	5,108	28.17 (4.48)	28.17 (4.48)	28.27 (4.48)	58.38 (4.48)

* Calculated F-value (Critical F-value for p = .001)

A significant difference at the .001 level occurs when $F_{critical} < F_{calculated}$.

TABLE 2

STATION CATEGORIES

SEA LEVEL	PLAINS	MOUNTAIN
Brownsville, Tx	Dayton, OH	Denver, CO.
Chatham, MA	Flint, MI	Great Falls, MN
Miami, FL	Glasgow, MT	Medford, OR
Portland, ME	Green Bay, WI	Salem, OR
Washington, DC	Greensboro, NC	Spokane, WA
Waycross, GA	International Falls, MN	Winslow, AZ
	North Platte, NE	
	Sault Ste. Marie, MI	
	Topeka, KS	

REFERENCES

1. Draper, N. R., and H. Smith, Applied Regression Analysis, John Wiley, New York, 1981.

2. Murphy, E.A., and K. G. Scharr, Modeling Turbulence in the Lower Atmosphere Using Richardson's Criterion, AFGL-TR-81-0349, ADA 115244, 1981.

3. Murphy, E. A., R. B. D'Agostino, and J. P. Noonan, Patterns in the Occurrences of Richardson Number Less Than Unity in the Lower Atmosphere, J. Appl. Meteorol., 21, 321-333, 1982.

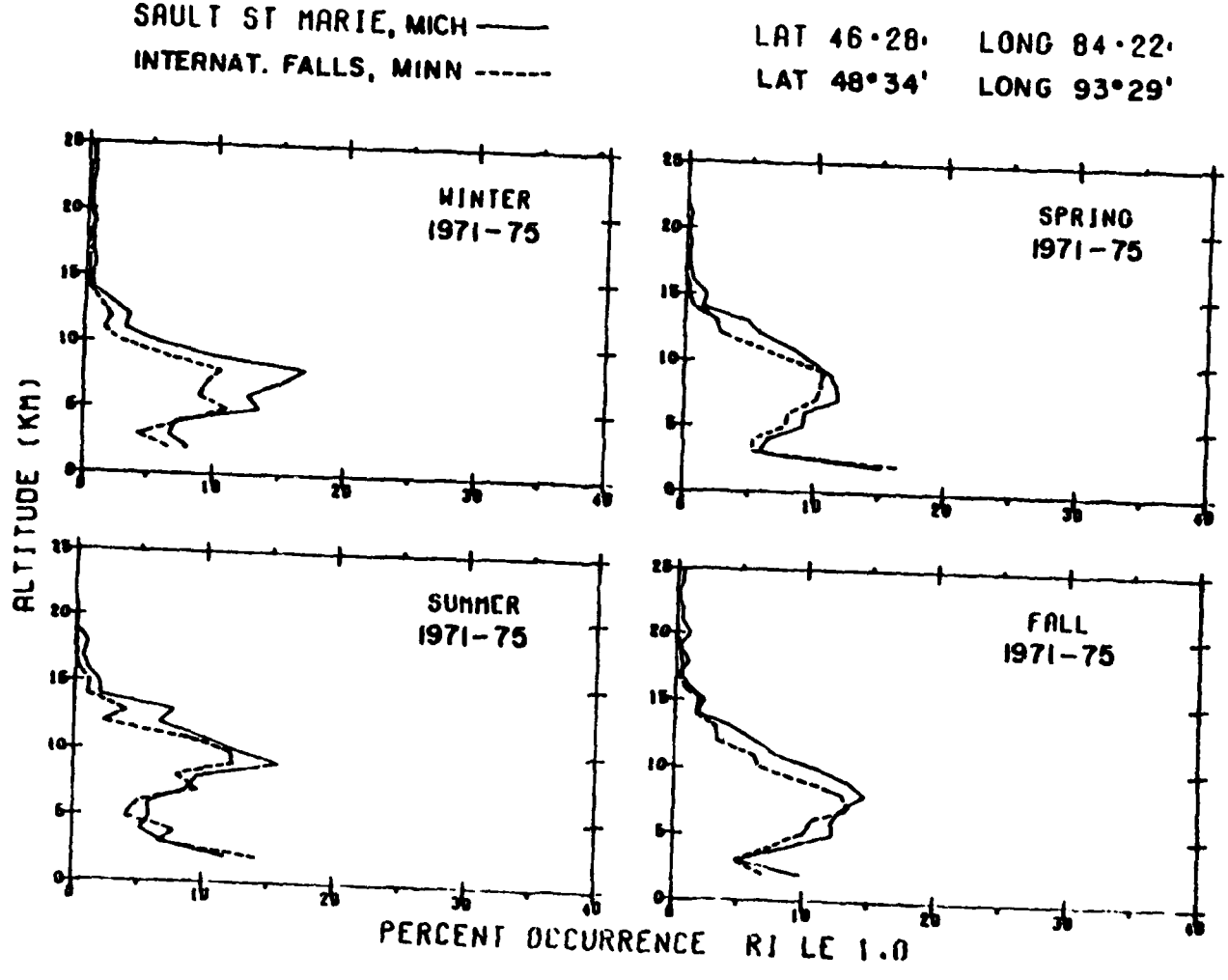

SAULT ST MARIE, MICH ——
INTERNAT. FALLS, MINN ------
LAT 46·28· LONG 84·22·
LAT 48°34' LONG 93°29'
WINTER
1971-75
SPRING
1971-75
SUMMER
1971-75
FALL
1971-75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig. 1a

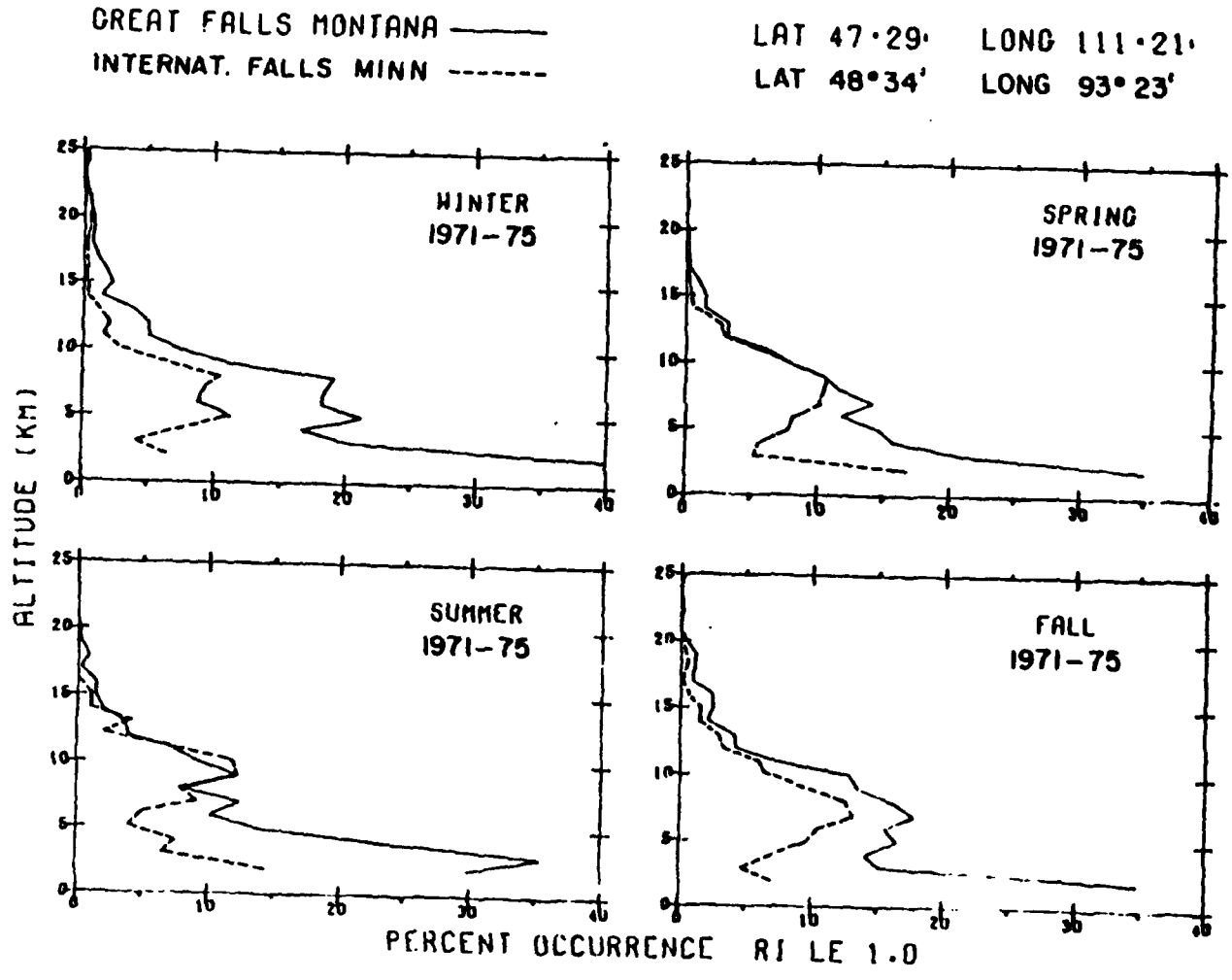

GREAT FALLS MONTANA
INTERNAT. FALLS MINN
LAT 47·29· LONG 111·21·
LAT 48°34' LONG 93°23'
WINTER
1971–75
SPRING
1971–75
SUMMER
1971–75
FALL
1971–75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig.1b

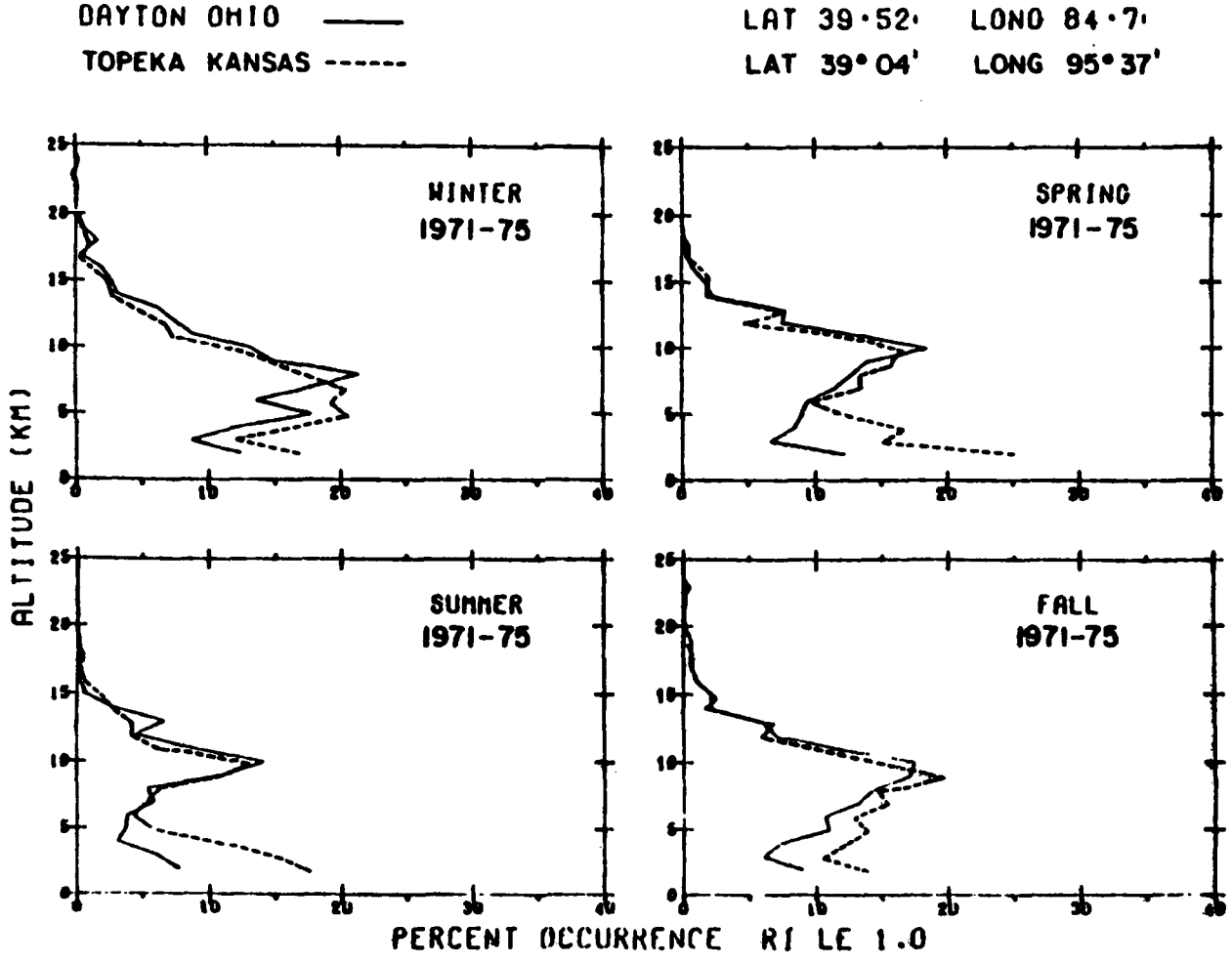

DAYTON OHIO
TOPEKA KANSAS
LAT 39·52·
LONG 84·7·
LAT 39° 04'
LONG 95° 37'
WINTER
1971-75
SPRING
1971-75
SUMMER
1971-75
FALL
1971-75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig. 2a

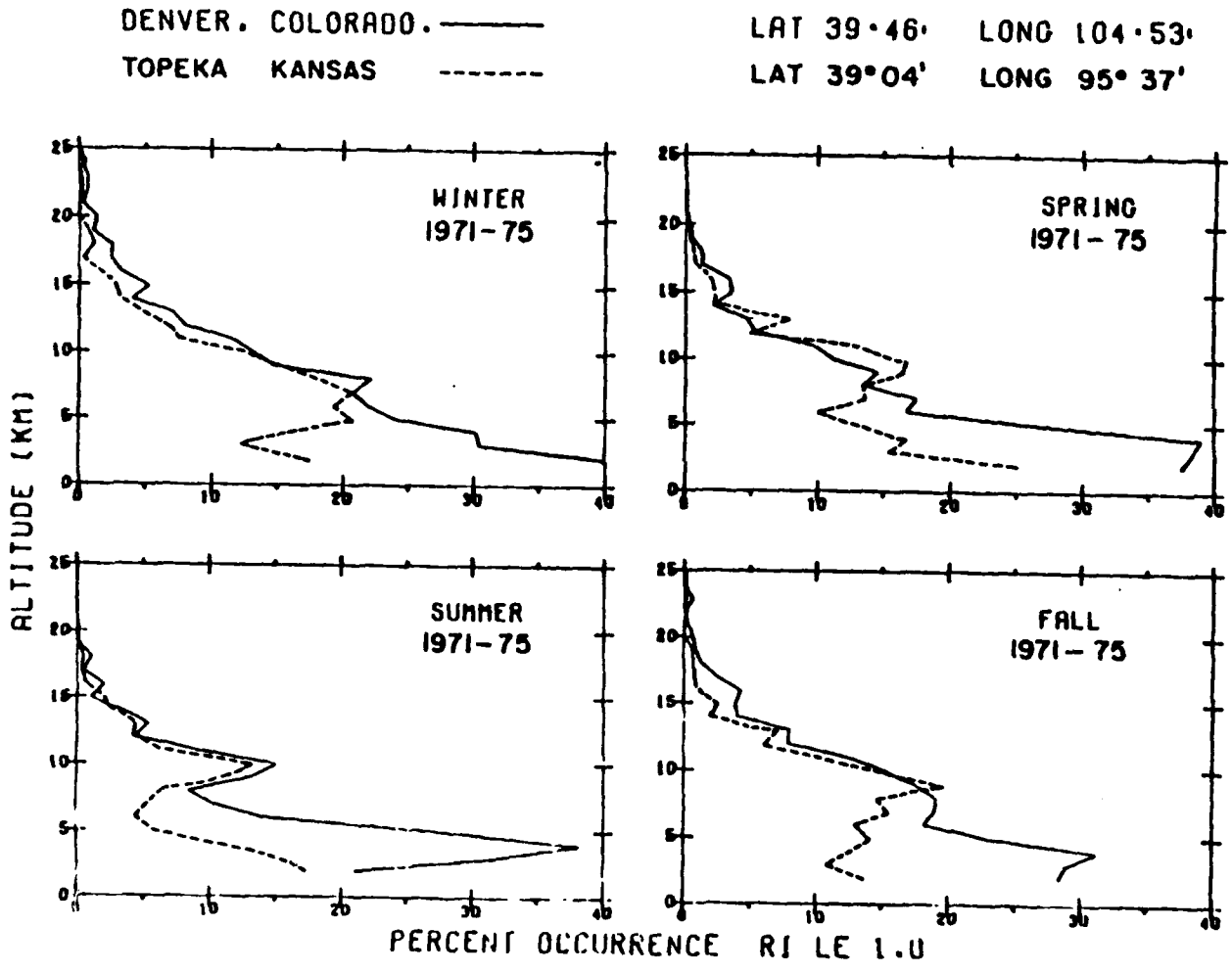

DENVER. COLORADO. ——— LAT 39·46· LONG 104·53·
TOPEKA KANSAS ------- LAT 39°04' LONG 95° 37'
WINTER
1971-75
SPRING
1971-75
SUMMER
1971-75
FALL
1971-75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig. 2b

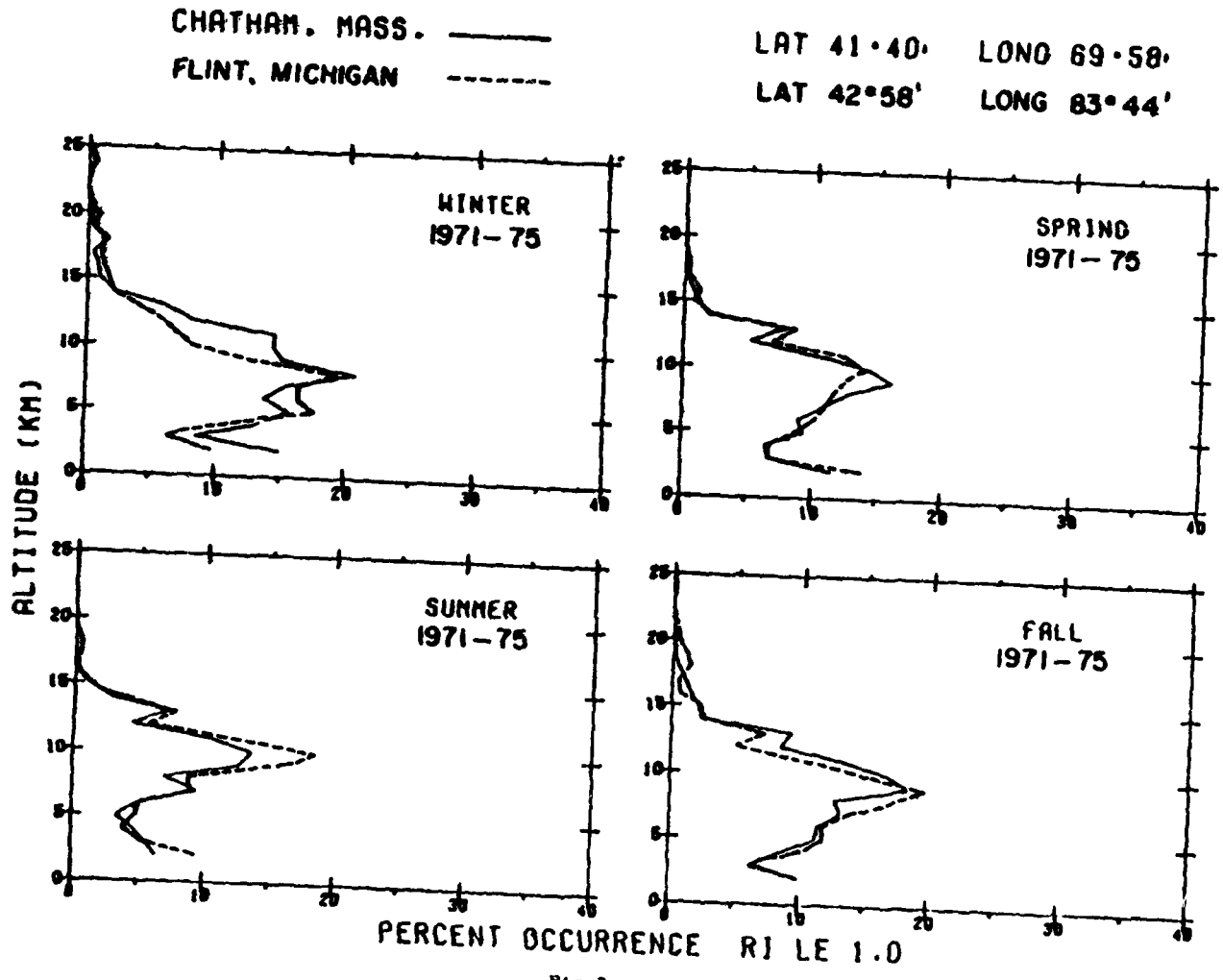

CHATHAM. MASS.
FLINT. MICHIGAN
LAT 41·40· LONG 69·58·
LAT 42°58' LONG 83°44'
WINTER
1971–75
SPRING
1971–75
SUMMER
1971–75
FALL
1971–75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig 3a

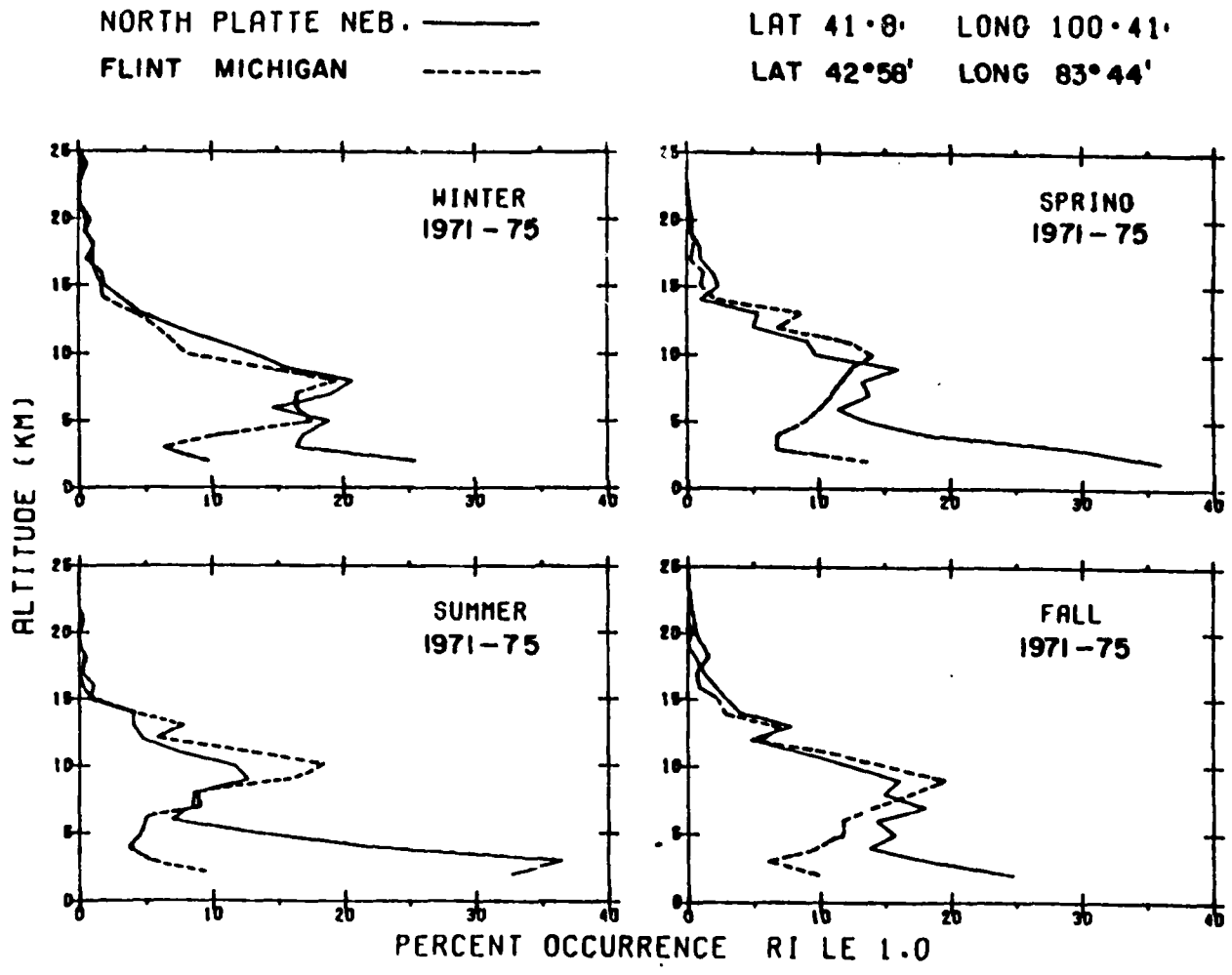
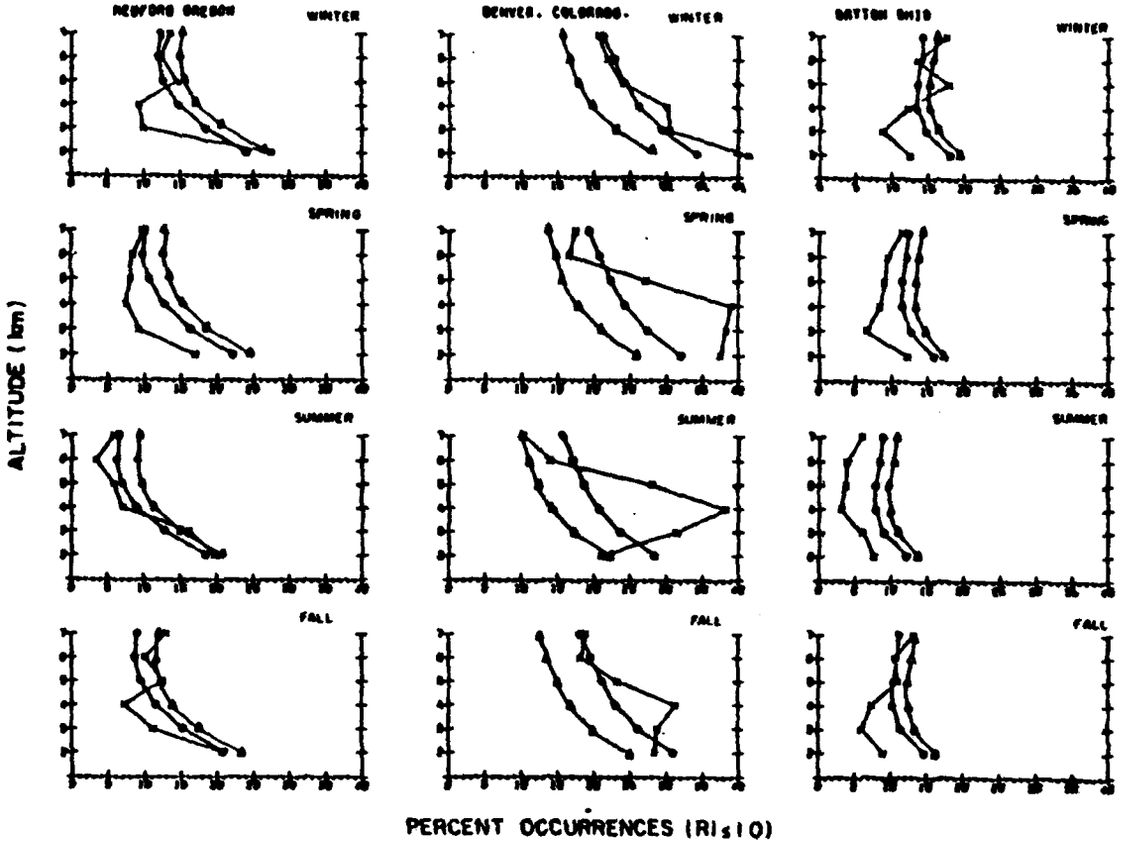

NORTH PLATTE NEB. ——— LAT 41·8· LONG 100·41·
FLINT MICHIGAN -------- LAT 42°58' LONG 83°44'
WINTER
1971 – 75
SPRING
1971 – 75
SUMMER
1971 – 75
FALL
1971 – 75
ALTITUDE (KM)
PERCENT OCCURRENCE RI LE 1.0

Fig. 3b



X – ACTUAL
O – PREDICTED WITH TERRAIN VARIABLE
Δ – PREDICTED WITHOUT TERRAIN VARIABLE

FIG. 4

A TECHNIQUE FOR ESTIMATING UNIFORM GRIDS OF ION DENSITY FROM IRREGULARLY-SPACED DATA

1. OVERVIEW OF THE SOFTWARE PACKAGE

The software package developed by Bedford Research Associates processes irregularly-spaced satellite measurements of ion density for the Air Force Global Weather Center (AFGWC) and extrapolates them to a regularly-spaced grid. The extrapolation procedure uses an estimation algorithm called the Best Linear Unbiased Estimator (BLUE). The package is completely self-contained, needs no outside algorithms, and needs no input parameters except the satellite data. A short overview of the structure of the software package follows. Details about the theory of BLUE and the practice of parameter estimation are found in subsequent sections.

Figure 1-1 shows the overall structure of the software, which can be divided into three basic modules as shown. The first module performs the actual reading and editing of the data file. The input to this module is a 24-hour set of GWC satellite data containing values of longitude, latitude and density. The data are subdivided into 6 AM and 6 PM data files, according to the local times along the satellite path at which the measurements are made. AM data occur when the satellite is travelling south to north, while PM data are acquired north to south. The grid to which the data is extrapolated covers 0^{o}-80^{o}N latitude with a delta of 15^{o}. Since the range of the GWC data is approximately 82^{o}S-82^{o}N, and 0^{o}-360^{o} longitude, all data not in the grid range are eliminated. Finally, the data files are further reduced by sampling every fourth observation of each file. This sampling is required to meet constraints imposed by memory and computer time considerations. The maximum number of points which can correctly be accepted by BLUE is 100; the sampling just described results in approximately ninety points per file.

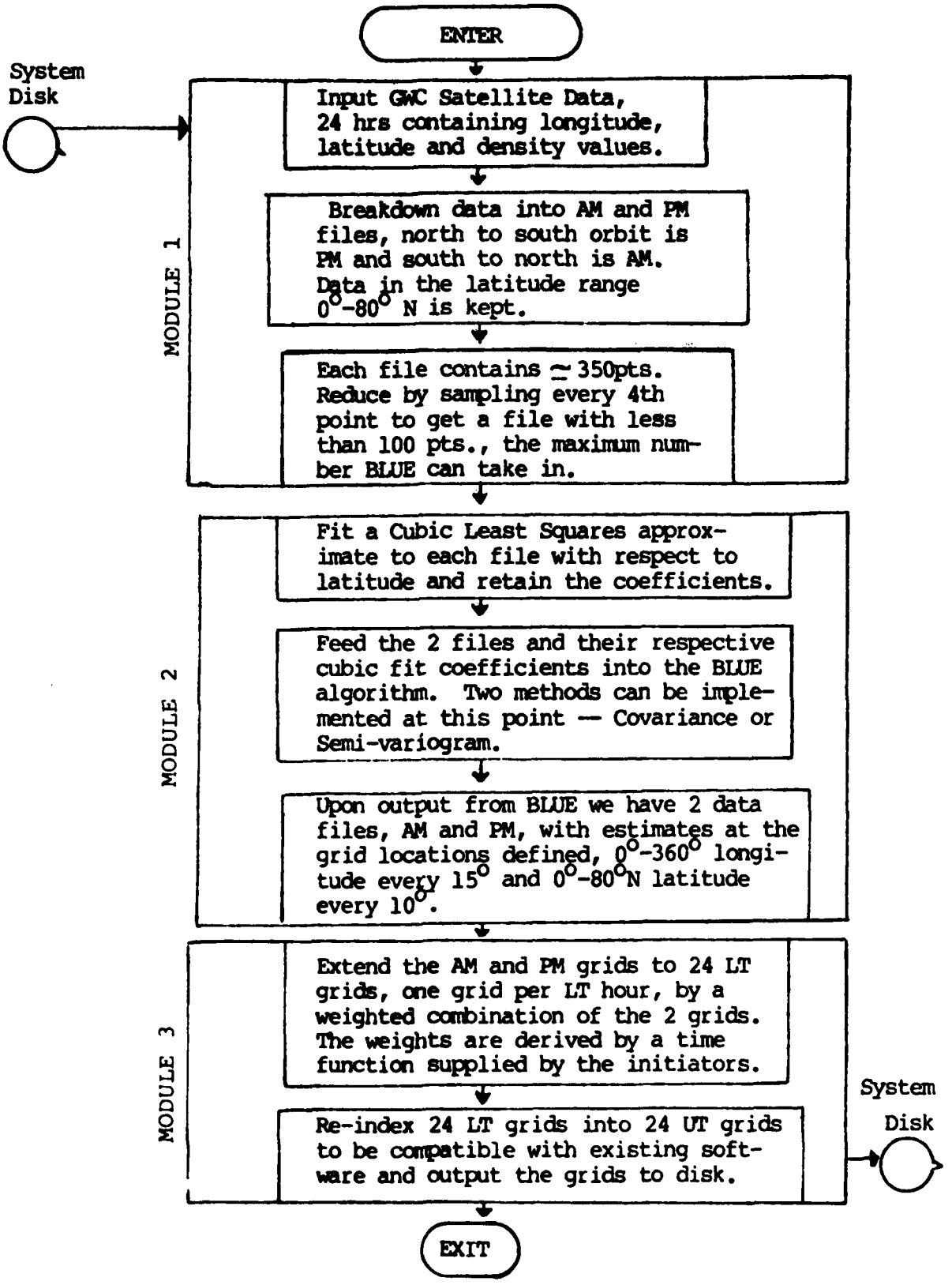


FIGURE 1.1: Software Package: Basic Structure

The second module executes the actual BLUE algorithm to extrapolate the reduced satellite readings to the 6AM and 6PM grids. Two versions of the second module have been developed, employing slightly different modelling assumptions. The first version, which is currently operational at AFGWC, assumes that a trend or mean structure in the ion density field can be subtracted from the observations prior to performing the extrapolation. The trend is restored to the estimated grid values at the end of the BLUE algorithm. The second version, currently under development, applies the BLUE algorithm to the actual observations without prior subtraction of the trend, but adds a constraint that ensures consistency between the mean of the observed values and the mean of the estimated grid values. In addition, the first version employs the covariance function to model the spatial correlation of the density field, while in the second version the semi-variogram function is being tested as a possible correlation model. The theory and modelling techniques used in the first version are presented in Section 2, while Section 3 discusses the theory and modelling being developed under the second version. In both versions, the final output of the second module is a uniform grid of unbiased, optimal estimates of ion density for each set of 6AM and 6PM observations.

In the last module, the 6AM and 6PM grid estimates are extended to a 24-hour period, providing estimated grids at one-hour increments of local time. A function defining the variation of density values with local time, shown in Figure 1-2, is used to derive a set of time-dependent weights. These weights are used to extend the 6AM and 6PM files to 24-hour local time files. The time function currently used was arrived at through discussion with the initiators. All points on the grid are assumed to vary equally with time. This relatively crude approximation reflects currently available knowledge and data on time variation, and refinement is planned as more information becomes available.

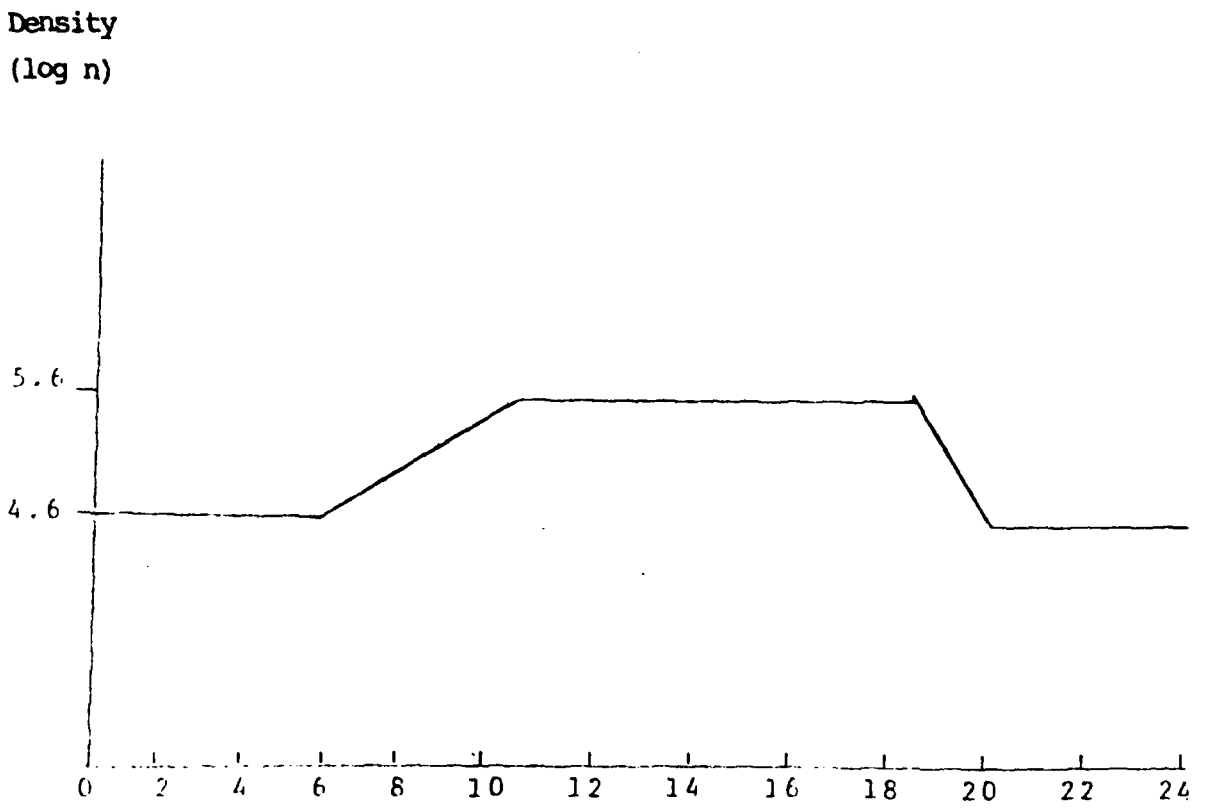


FIGURE 1-2

Initially Assumed Function of Density vs. Local Time

2. MODEL ASSUMPTIONS AND THEORY: BLUE VERSION 1

As noted in Section 1, the version of the software currently operational at AFGWC varies somewhat from the routines now under development for subsequent implementation at AFGWC. The difference lies in the second module which performs the actual estimation of the density field at grid locations. In both versions, the trend or mean structure of the density field is assumed to vary with latitude as a cubic polynomial.

mean: $$m(\underline{u}) = E\,[Z(\underline{u})] = \beta_o + \beta_1 Y + \beta_2 Y^2 + \beta_3 Y^3$$

where $E\,[.]$ is the expectation operator

$Z(\underline{u})$ is the ion density field

β_o, β_1, β_2 β_3 are the cubic polynomial coefficients

In Version 1, this cubic trend is subtracted from the observations and the residual field thus obtained used to estimate the grid values. The trend is restored to the estimates at the end of the BLUE algorithm. The second version leaves the trend in the observations while imposing a constraint to ensure consistency in the mean of the final estimated values. Section 2.1 presents the model assumptions and theory underlying the BLUE algorithm as implemented in Version 1, and Section 2.2 discusses the correlation model fitting techniques. The corresponding topics for Version 2 are discussed in Section 3.

2.1 BLUE ASSUMING STATIONARY MEAN

In the original implementation the BLUE algorithm was performed with the residual field

$$Y(\underline{u}) = Z(\underline{u}) - m(\underline{u})$$

The known second moment characteristic of $Y(\underline{u})$ are

mean: $E[Y(\underline{u})] = 0$

covariance: $Cov(\underline{u}_1, \underline{u}_2) = Cov(h) = E[Y(\underline{u}_1) Y(\underline{u}_2)]$

where

$$h = \left| \underline{u}_1 - \underline{u}_2 \right|$$

The linear estimator of a residual value $Y(\underline{u}_o)$ on the grid is:

$$\hat{Y}(\underline{u}_o) = \sum_{i=1}^{N} \lambda_i Y(\underline{u}_i)$$

given the mean-subtracted observation set, $Y(\underline{u}_i)$, $i=1,2,\ldots N$

Unbiasedness requires that

$$E[\hat{Y}(\underline{u}_o)] = E[Y(\underline{u}_o)]$$

which reduces to the condition:

$$\sum_{i=1}^{N} \lambda_i = 1$$

The estimation variance is

$$\sigma_e^2 = \text{Var}\,[Y(u_o) - Y(u_o)\,] = \text{Var}\,[\sum_{i=1}^{N} \lambda_i\, Y(u_i) - Y(u_o)\,]$$

$$= \text{Cov}\,(0) - 2 \sum_{i=1}^{N} \lambda_i\, \text{Cov}\,(h_{io})$$

$$+ \sum_{i=1}^{N} \sum_{i=1}^{N} \lambda_i \lambda_j\, \text{Cov}\,(h_{ij})$$

Minimizing the above subject to the condition $\sum_{i=1}^{N} \lambda_i = 1$ yields the following set of N+1 linear equations in the λ_i's and the Lagrange multiplier, μ .

$$\sum_{j=1}^{N} \lambda_j\, \text{Cov}\,(h_{ij}) + \mu = \text{Cov}\,(h_{io}) \qquad i=1,2\ldots N$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

Rewriting in Matrix notation

$$K \quad \underline{\lambda} \quad = \quad \underline{C}$$

where

$$\left[\begin{array}{cccc} \overbrace{\qquad\qquad\qquad\qquad}^{N+1} & & & \\ \text{Cov}\,(h_{11}) & \text{Cov}\,(h_{12}) & \ldots\ \text{Cov}\,(h_{1N}) & 1 \\ \text{Cov}\,(h_{21}) & & & \\ \vdots & & & \\ \text{Cov}\,(h_{N1}) & \ldots\ \text{Cov}\,(h_{NN}) & & 1 \\ 1 & 1 & 1 & 0 \end{array}\right] \Bigg\} N+1$$

$$\underline{\Lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \lambda_N \\ 1 \end{bmatrix} \qquad \underline{C} = \begin{bmatrix} \text{Cov } (h_{10}) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \text{Cov } (h_{N0}) \\ 1 \end{bmatrix}$$

K is symmetric (N+1) X (N+1) matrix and the solution requires its inversion yielding

$$\underline{\Lambda} = K^{-1} \underline{C}$$

The BLUE of $Z(\underline{u}_o)$ can hence be formed from the optimal set of λ_i's. Also, the optimal estimation variance is

$$\sigma_e^2 = \text{Cov } (0) - \sum_{i=1}^{N} \lambda_i \text{ Cov } (h_{jo}) - \mu$$

Note that the matrix K need be computed and inverted only once for all points on the grid to be estimated, since the same set of observations is used in estimating the whole grid. Only the vector $\underline{C}$ varies from one grid value to the next.

2.2 MODEL PARAMETER ESTIMATION: ESTIMATING THE EXPONENTIAL COVARIANCE

The BLUE algorithm currently implemented on the UNIVAC computer at AFGWC performs a minimum variance estimation using a linear combination of known observations to produce unbiased estimates of the unknown grid locations. This estimation procedure requires prior knowledge of the spatial correlation of the ion density data. This version of BLUE models spatial correlation in terms of the exponential covariance function

$$\text{Cov } (\underline{u}_1, \underline{u}_2) = \sigma^2 \exp (-ah)$$

where

$\underline{u}_1$, $\underline{u}_2$ are the coordinates of the two points under consideration

$h = \left| \underline{u}_1 - \underline{u}_2 \right|$

a = exponential parameter

σ^2 = point variance of the ion density field

Implementation of the BLUE algorithm then requires the prior estimation of a and the point variance σ^2.

The point variance σ^2 can be easily estimated on-line from each set of 6AM and 6PM observations as they become available. The value a was estimated from the data available during model development. This value was arrived at through trial and error and a procedure whereby each observed value is treated as an unknown point and estimated using the others. Statistics formed from the difference between the estimated and actual observed values are used as measures of the goodness of the assumed value of a. Trial and error show that the value a = 0.2 yielded the best values for these statistics and this value has been incorporated into the software now operational at AFGWC.

It is, of course, to be expected that each set of 6AM or 6PM observations may yield a different correlation structure. The new version of BLUE to be implemented will incorporate procedures to estimate on-line the parameters of the correlation function (the semivariogram in the new version) for each set of 6AM or 6PM observations.

3. MODEL ASSUMPTIONS AND THEORY: BLUE VERSION 2

To take care of variation in the spatial correlation structure of the density field from one 12-hour observation period to the next, on-line estimation of correlation parameters for each set of 6AM or 6PM observations was deemed necessary. Preliminary analysis with available data also points to the advantage of using the semivariogram (defined below) to measure spatial correlation. In addition, theoretical considerations as well as evidence of increased precision in practice, suggest that the observation set should be used directly (i.e. without subtracting the mean) to estimate the unobserved grid values. The above changes are being incorporated in a new version of the BLUE algorithm (the second module of the software package) and analysis of their validity is in progress. Section 3.1 presents the BLUE algorithm using the actual observation set (rather than the residual set) while Section 3.2 discusses model fitting techniques for the semivariogram currently under development.

3.1 BLUE ASSUMING KNOWN NON-STATIONARY MEAN

The ion density field $Z(\underline{u})$ is assumed to be a random process in $\mathcal{R}^2$ with known second-moment characteristics. The mean expressed as

$$E\,[Z(\underline{u})] = m(\underline{u})$$

is known for all points in the area of interest (i.e., all points on the grid where values are to be estimated), and in this case a cubic polynomial varying with latitude only. The semivariogram defined as

$$\gamma(h) = \frac{1}{2} \text{Var}\,[Z(\underline{u}_1) - Z(\underline{u}_2)]$$

where

$$h = \left| \underline{u}_1 - \underline{u}_2 \right|$$

is also known. In our case, the semivariogram is isotropic and spherical in structure with parameters that may be estimated (see Section 3.2).

Given a set of observations $Z(\underline{u}_i)$, i=1,...N, it is desired to estimate values on an unobserved regular grid. Let $Z(\underline{u}_o)$ be a point on the grid.

The linear estimator of $Z(\underline{u}_o)$ is formed as follows:

$$\hat{Z}(\underline{u}_o) = \sum_{i=1}^{N} \lambda_i Z(\underline{u}_i)$$

Unbiasedness is maintained by requiring

$$\hat{E}[Z(\underline{u}_o)] = E[Z(\underline{u}_o)]$$

which is re-expressible in terms of the known means as

$$\sum_{i=1}^{N} \lambda_i m(\underline{u}_i) = m(\underline{u}_o)$$

The estimation variance is expressed as

$$\text{Var}[\hat{Z}(\underline{u}_o) - Z(\underline{u}_o)]$$

$$= \text{Var}\left[\sum_{i=1}^{N} \lambda_i Z(\underline{u}_i) - Z(\underline{u}_o)\right]$$

Matheron (1970) has shown that the above variance is finite only if the following condition on the weights, λ_i, holds:

$$\sum_{i=1}^{N} \lambda_i = 1$$

Given the above condition, the estimation variance can then be expressed in terms of semivariograms as

$$\text{Var} \left[\sum_{i=1}^{N} \lambda_i Z(u_i) - Z(u_o) \right]$$

$$= - \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma(h_{ij}) + 2 \sum_{i=1}^{N} \lambda_i \gamma(h_{io})$$

The "best" criterion requires that this esimation variance be minimized subject to the two constraints on the λ_i. A Lagrange multiplier formulation yields the following set of N+2 linear equations in the λ_i's and the multipliers, μ_o and μ_1.

$$\sum_{j=1}^{N} \lambda_j \gamma(u_j - u_i) + \mu_1 + \mu_o m(u_i) = \gamma(u_i - u_o)$$

$$i = 1,2,\ldots N$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

$$\sum_{i=1}^{N} \lambda_i m(u_i) = m(u_o)$$

Or in matrix formulation

$$G \underline{\Lambda} = - \underline{W}$$

where

$$G = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots\cdots & \gamma(h_{1N}) & m(\mu_1) & 1 \\ & \ddots & & & & \\ \gamma(h_{21}) & & \ddots & \gamma(h_{2N}) & m(\mu_2) & 1 \\ \vdots & & & \vdots & \vdots & \vdots \\ \gamma(h_{N1}) & \cdots\cdots & & \gamma(h_{NN}) & m(\mu_N) & 1 \\ m(\mu_1) & \cdots\cdots & & m(\mu_N) & 0 & 0 \\ 1 & \cdots\cdots & & 1 & 0 & 0 \end{bmatrix} \quad \left.\begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix}\right\} N+2$$

$$\underline{\Lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu_o \\ \mu_1 \end{bmatrix} \left.\begin{matrix} \\ \\ \\ \\ \\ \end{matrix}\right\} N+2 \qquad \underline{W} = \begin{bmatrix} \gamma(h_{1o}) \\ \vdots \\ \gamma(h_{1N}) \\ m(\mu_o) \\ 1 \end{bmatrix}$$

The solution then becomes

$$\Lambda = G^{-1} \underline{W}$$

and the BLUE of $Z(\underline{u}_o)$ can then be formed from the optimal set of λ_i, i=1,2...N.

In addition, the minimum estimation variance is:

$$\sigma_e^2 = \sum_{i=1}^{n} \lambda_i \gamma (\underline{u}_i - \underline{u}_o) + \mu_1 + \mu_o m(\underline{u}_o)$$

The above algorithm is implemented for each point $\underline{u}_o$ on the regular grid. Since the same observations $Z(\underline{u}_i)$, i=1,...N are used for estimating every point on the grid, it is apparent that the matrix G needs to be inverted only once and only the R.H.S. vector $\underline{W}$ varies from point to point on the grid yielding a different set of optimal λ_i's for each grid to be estimated.

3.2 MODEL PARAMETER ESTIMATION: ESTIMATING THE SEMIVARIOGRAM

The BLUE algorithm requires the specification of a model for the mean structure (or trend) of the ion density field and a correlation structure in terms of either a covariance or semivariogram. In the new version of the software, a cubic polynomial which is a function only of the latitude is assumed for the trend and the form of the spatial correlation is specified in terms of an isotropic spherical semivariogram. The respective equations are:

mean:

$$m(\underline{u}) = E[Z(\underline{u})]$$

$$m(\underline{u}) = \beta_o + \beta_1 y + \beta_2 y^2 + \beta_3 y^3$$

semivariogram:

$$\gamma(h) = (\underline{u}_1 - \underline{u}_2])$$

$$= \frac{1}{2} \text{Var}[Z(\underline{u}_1) - Z(\underline{u}_2)]$$

$$= \frac{1}{2} E[(Y(\underline{u}_1) - Y(\underline{u}_2))^2]$$

spherical semivariogram:

$$\gamma(h) = \begin{cases} C_o + C & h \geq a \\ C_o + C\left[1.5\left(h/a\right) - 0.5\left(h/a\right)^3\right] & h < a \\ 0 & h = 0 \end{cases}$$

where $Z(\underline{u})$ is the ion density field

$Y(\underline{u})$ is the residual field

$$Y(\underline{u}) = Z(\underline{u}) - m(\underline{u})$$

and

$(\underline{u}) = (x, y)^T$ are the geographical coordinates

h is a distance measure computed as if the points $(\underline{u}_i)$ lie on a cylinder (instead of a sphere). This model assumption has the effect that observation points near the North Pole which lie on different orbits are pushed apart so that the effective correlation across orbits is smaller than would otherwise be the case if actual distances on a sphere are assumed.

β_0, β_1, β_2, β_3 = coefficients of the cubic trend

Co, C, a = parameters of the spherical semivariogram

a is known as the range

Co the nugget effect

and Co + C the sill.

These parameters are illustrated geometrically in Figure 3-1.

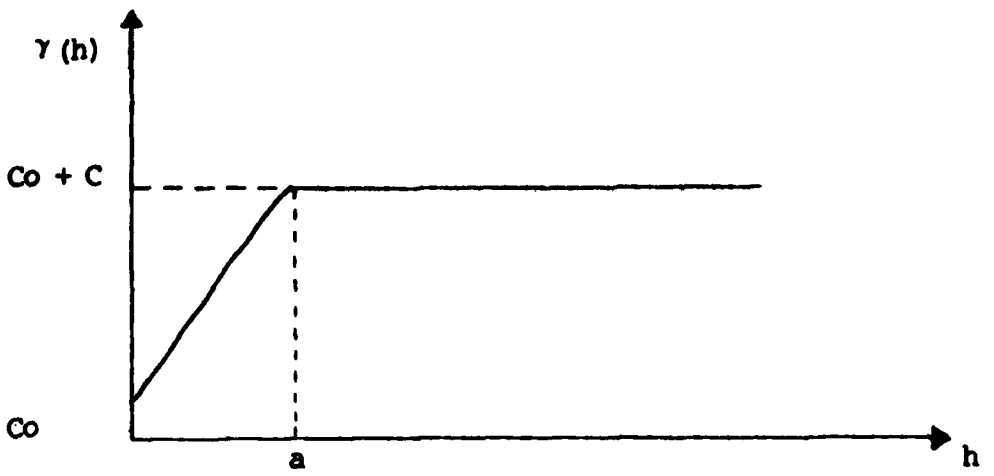


Figure 3-1: Spherical Semivariogram

The model assumption of a cubic polynomial varying with latitude for the mean structure was suggested by ion density data currently available and confirmed by AFGWC. Tests with available data also sugest that the form of the spherical semivariogram adequately models the residual correlation. Many physical phenomena in $\mathcal{R}^2$ are also known to exhibit the spherical semivariogram structure (see David [1977] and Chua [1980]) and the parameters are amenable to automatic parameter fitting procedures.

Although the form of the trend and semivariogram is assumed not to vary from one 12-hour period (6AM or 6PM) to the next, the parameters βi (i=0,...3) and Co, C, and a will be estimated for every set of 6AM and 6PM observations.

The coefficients βi (i=0,...3) will be estimated using a simple least squares regression.

With the trend subtracted from the observations, the residuals are now used to estimate the semivariogram parameters. The procedure implemented here is a modification of the method suggested by David (1977).

An experimental or "raw" semivariogram is first computed from the set of 6AM (or 6PM) observations under consideration. Since the observations are irregularly distributed with respect to the distance measure h, the estimator of $\gamma(h)$ is the following:

$$\hat{\gamma}(h) = \frac{1}{2n_h} \sum_{i=1}^{n_h} \left\{ Y(u_i + h) - Y(u_i) \right\}^2$$

where pairs of observations are grouped by discrete distance intervals, say n_h pairs between h and $h+\Delta h$ and the value h in the above equation is an average value

$$h = \frac{1}{n_h} \sum_{i=1}^{n_h} h_i$$

The h value to be used in the current implementation is $\Delta h = 5.0$ which has shown to provide sufficient discretization to define the raw semivariogram structure.

To fit the spherical semivariogram, the sill is first established from an estimate of the variance of the field

$$Co + C = \sigma_Z^2$$

A weighted least squares regression is then performed to fit the rising part of the semivariogram. n_a points on the experimental semivariogram are selected for this regression where the $(n_a - 3)^{th}$ point is the first raw semivariogram value to exceed the sill value, Co + C. The regression will yield the value of the nugget effect Co and the intersection of the regression line and the sill Co + C defines the point

$$h = \frac{2}{3} a.$$

The estimated parameters are the input to the BLUE algorithm which estimates a regular grid of density values based on the estimated parameters and assumed model for trend and semivariogram.

3.3 OPTIONAL MODEL VALIDITY CHECK

Note that software is also available to test the validity and accuracy of the assumed models and the estimated parameters. Since the modelling assumptions of the form of the trend and semivariogram were made on the basis of the data currently available to the analysts, the uncertainty obviously exists whether the same assumptions will apply to the on-line data to be collected. To test these assumptions, a program is available which takes each set of 6AM or 6PM observations and the model parameters estimated from them and, by a process of estimating each observation using the others, establishes a set of statistics that measure the validity of the model and model parameters used. This program can be executed if a posteriori checks on model assumptions and re-adjustment of parameters are deemed necessary.

REFERENCES

1. Chua, S. H. and R. L. Bras, Estimation of Stationary and Non-stationary Random Fields: Kriging in the Analysis of Orographic Precipitation, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Report No. 255, Cambridge, MA, 1980.

2. David, M., Geostatistical Ore Reserve Estimation, Elsevier Scientific Publishing Co., New York, 1977.

3. Matheron, G., The Theory of Regionalized Variables and Its Application, Cashiers du Centre de Morphologie Mathematique, 5, Ecole des Mines, Fontainebleau, France, 1970.

LOW ENERGY ELECTRON AND PHOTON CROSS SECTIONS FOR O, N_2, O_2 AND RELATED DATA

A set of low energy (0-300 eV) cross sections of functions of energy and related date were compiled for use in calculating the photoelectron flux in the daytime ionosphere of the earth.

The upper atmosphere is composed of atoms, simple molecules, and their ions. When the sun is relatively quiet, the atmosphere can be described and modeled as being in the steady state. Sample data on the neutral particle composition of the upper atmosphere and ionizing electromagnetic radiation from the sun were included (e.g., values for neutral, ion, and electron densities and the associated temperatures found in the mid-latitude daytime ionosphere, and tables of the energy levels of the main species in the upper atmosphere).

The photoionization and photodissociation of the atmosphere species are the means of putting additional energy into the upper atmosphere. This is done by the extreme ultraviolet (EUV) region of the spectrum. The EUV spectrum is composed primarily of lines and, when examined in great detail, is highly structured. It must be measured above the atmosphere of the earth. Two sample measured solar spectra were included.

The ionospheric density is low enough for simple photoabsorption to result in re-emission and the total process resembles elastic scattering. The main interest is in the energetic electrons produced through photoionization. In the energy range in which photoionization is possible, absorption must also be considered. Photoionization and the photoabsorption cross sections for O, N_2, and O_2 were chosen for wavelengths from 34 to 1030A.

The recombination rates of atmospheric species are measured directly in laboratory experiments or determined from measured cross sections. The analysis of atmospheric data provides a check on the rates and assumptions involved. The cross sections for dissociative recombination are well behaved near .1 eV. Using this as the main part of the cross sections allowed the rates to be calculated.

The experimental total cross sections for ionization of O, N_2, and O_2 were reviewed. One difficulty in the use of these is that the energy of the other electron is undetermined since the ionic state and its excitation energy are unknown.

The electron-neutral particle cross sections must be known in order to calculate the electron distribution function for the Earth's ionosphere. The variety of states and cascading processes make experimental identification of the many cross sections difficult, especially for the higher lying states. As the energy increases, many states may be summed over: rotational states for vibrational excitation; vibrational (and rotational) states for electronic excitation; rotational (and possibly vibrational) virtual states for resonances. At energies high enough to allow the excitation of several different levels, the total cross section must be broken down into the sum of the separate cross sections. This is a difficult task and may cause ambiguities in the final result.

At the very low energies, measurements of the cross sections are difficult and much data comes from calculations. Due to the complexity of multi-electron molecules, approximations in the calculations must be used.

The differential cross sections have been measured for a number of the scattering interactions at specific energies. Since few different energy values were taken, in general, the report did not consider the angular dependence of cross sections except the implied dependence in the momentum transfer cross sections.

The momentum transfer (or diffusion) cross section is one of the series of integral cross sections which are weighed by a function of the scattering angle arising in an angular decomposition of the Boltzmann collision integral. It can be obtained by a weighted integral of the differential elastic cross section obtained from experiment or theory. Some experiments obtain the momentum transfer cross section directly, e.g. swarm experiments.

The cross sections for transfers between very low-lying states, such as rotational states, are hard to determine either experimentally or theoretically.

The ground state of oxygen has a fine structure of three levels being at energies of 0, .019, and .029 eV or of 0, 156, and 235° K. No direct experimental data is available. Comparison of calculations showed agreement.

The separation of successive vibrational energy levels is usually in the vicinity of .2 eV or 2000° K. In the ionosphere this is high enough so that the excited levels are relatively depopulated but low enough so that excitation by energetic electrons provides a sink for the electron energy, especially in the case of N_2. There have been many calculations of the separate vibrational excitation cross sections. The calculation chosen agrees well with the experimental general structure of the separate cross sections and is about 10% larger than the measured absolute total cross section.

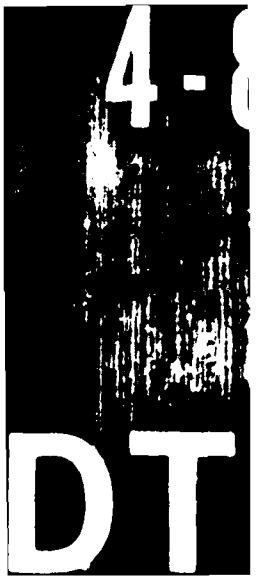
The electronic states of the atmospheric species start at an energy of a few electron volts. These states are not thermally excited. They do form a sink for the energy of the photoelectrons through excitation by the electrons and subsequent radiative decay. It is at these energies that the separation of the effects from different states becomes difficult. As a result, the determination of a particular cross section at a particular electron energy depends on knowing the others or on having a method of separating them. Comparing sets of cross sections with other data which can be connected with them (e.g. electron transport coefficients) allows

determination of a set which is self-consistent in a given context. If the cross sections are to be used to model a similar situation, such a set should give good results if it is sufficiently complete. Such sets were determined using electron transport coefficients for O_2 and N_2. These sets served as a guide in making a choice between various experimental cross sections.

Atomic oxygen is difficult to handle experimentally because it is so reactive. The theoretical approach is also difficult due to the need to properly represent the open shell configuration of the atom, its interactions with the incoming electron including exchange and polarization, and the effects of low-lying states of the atom and its negative ion, i.e., excitation and resonances.

The cross sections considered above described the collision of an electron and an atom or molecule in which the electron loses some of its energy. The opposite process, in which the electron gains energy through excitation of the atom or molecule, was related to the first.

Molecular dissociation by electron impact can be considered as a single complex process. The total cross section does not determine the electron energy in the three particle final states and more structure must be included either experimentally or theoretically. It is natural to divide the process so that the significant parameters are excitation cross sections for the different molecular states and branching ratios. Below the lowest dissociation energy the states cannot dissociate; above it, dissociation may be probable or not depending on the relative accessibility of dissociated and nondissociated states. Rydberg states, which could not ionize, were considered to dissociate, as were all the O_2 states for which dissociation was energetically possible.


4-
DT